

Energy shame

The history of energy research highlights the importance and inadequacies of markets, and a yawning gap in the priorities of governments. It's time for a radical change.

Frances Cairncross, chair of Britain's Economic and Social Research Council, has been thinking about the economics of climate change longer than most natural scientists and economists. In her presidential address to the annual meeting of the British Association for the Advancement of Science this week, she rightly emphasized one of the most important things that governments can do: invigorate and focus research into the basic and translational science needed for new energy-conversion technologies.

Solar power is a case in point. Its great economic attraction is that, unlike nuclear power or carbon capture and storage, it does not need vast capital investment in order to spread. Its products just need to be priced in such a way that consumers and companies want to buy them. Once that point is reached, a solar-cell factory can produce the capacity to generate electricity as easily as a power station does, thus offering the possibility of exponential growth.

As we report on page 19, a boom in the solar-energy business, led by Japan and Germany, has now attracted serious interest from, among others, the technologists and venture capitalists of California's Silicon Valley. The people who brought the world Moore's law are eager to help it sustain itself through clean technology while accumulating yet more wealth on the way. Its most vigorous proponents suggest that attracting the attention of high-tech entrepreneurs could in itself be an end to our energy woes — a “distributed Manhattan project that attracts the smartest, most ideal people for the task”, as Bill Gross, serial entrepreneur and trustee of the California Institute of Technology, put it to *The New York Times* earlier this year.

But a healthy respect for the power of entrepreneurs and free markets cannot hide the fact that they do best when choosing between possibilities that are close to market, rather than inventing entirely new options. There is research into new materials and technologies that small start-up companies can't do, and that larger, more staid ones, if history is a guide, won't.

History may not be a guide, of course. But even the possibility that

the research may not get done is a reason for government to step in and ensure that it does, while trying not to crowd out private capital in the process. The current solar boom is dependent on old and trusted technologies — the companies now piling in are mostly finding new ways to manufacture and process familiar products. If the current rate of heady growth is to keep going for the quarter-century needed to start making a real change in the world's energy outlook, we will need new materials to capture the Sun's power ever more cheaply and easily, and new solutions to the problem of storing it.

Many scientists are eager to set out in search of those technologies. It is essential that curiosity-led research should flourish in these areas, and that funding bodies should encourage it so to do.

There are also areas where directed research might come into its own — where the best approach may be to try out lots of possibilities, rather than go with what a few bright people think is best. One of the strengths of the Manhattan project was that it tried out as many roads to nuclear weaponry as seemed plausible. Governments need to be willing to take advice on directed research and then make it happen, rather than just hoping that curiosity will triumph unaided.

Talk of a Manhattan project to tackle the generation and storage of ‘clean’ energy may seem overblown. It shouldn't. The challenge of increasing energy use in the developing world while at least stabilizing and ideally decreasing carbon dioxide emissions is immense. It is to the abiding shame of the world's governments that, as the threat to the climate has become ever more apparent over the past two decades, funding for energy research and development has actually fallen. To suggest that spending on energy research should be limited only by the capacity of scientists and technologists to make practical use of it is not to be profligate, but rational. ■

“It is to the abiding shame of the world's governments that over the past two decades, funding for energy research has actually fallen.”

Beauties of synthesis

How to excite an organic chemist.

Many organic chemists spend their days searching for creative solutions to real-world problems, yet the media pays them just a fraction of the attention devoted to physicists or biologists. Even fellow scientists think organic chemistry is esoteric.

Ask a group of organic chemists why they love their work, however, and most will tell you that it enables them to make things that no one else has made before. Many spend their days trying to make large, architecturally complex molecules, and some believe that the

journey taken to prepare the compound is as important as arriving at the destination. Indeed, chemists are frequently drawn to the field because there is not just one way to solve the problem, and the search can reveal a bit more about how the world works.

When a complex synthesis is first reported at a conference, the excitement can be palpable. The audience can be inspired by the speaker's innovative approach, but there may be another reason for their exhilaration: they might be able to apply the lessons learned from that synthesis to their own research.

An impressive synthesis may be described as ‘beautiful’ or ‘elegant’ because of an aesthetic appreciation of the molecule itself or the way it was created. But it often implies that the approach was creative and that the molecule was made efficiently. Because chemists design their



syntheses by looking at the target molecule and working backwards towards simpler starting materials, chemists can impress their peers by pursuing strategies no one else noticed, by using original sequences of previously disparate reactions, or by developing new reactions that enable them to construct bonds in an unexpected way.

The work isn't over when the first synthesis of a complex molecule is reported — there is always plenty of room for improvement, and chemists are fired up by the desire to make things better. Second-generation syntheses are often very different from the original, and may have fewer steps or contain new and unexpected chemical reactions. This iterative process is not just an academic exercise: process chemists in pharmaceutical companies are often the unsung heroes of the synthetic world, as they devise incredibly effective syntheses to produce kilogram quantities of potential drugs. By doing so, they save millions of dollars for their companies, and can reduce the environmental impact of the synthesis by minimizing the amounts of waste produced or the quantities of solvents required.

In a similar vein, chemists enjoy finding improved catalysts. Catalysts exist for many reactions but are often expensive, toxic or impractical for anything other than simple molecules. By exploring the chemical mechanism of the catalyst, or perhaps by just plain luck, chemists can develop second- and third-generation catalysts that dramatically outperform the first-generation system. This can be invaluable both in the lab and for improving industrial processes.

A common misconception is that organic chemists now have a

complete 'toolbox' of reactions that can be applied to any synthesis. But the toolbox is certainly not full; many desirable reactions remain elusive. This is especially true in the area known as 'asymmetric catalysis', which involves the creation of a chiral material from a non-chiral substrate — molecules can exist in left-hand and right-hand forms, and the development of catalysts that can selectively make one of the two forms is a major challenge in the field. To outsiders, this is a particularly obscure field of work, but the inherent challenges involved attract vast numbers of researchers. Furthermore, the products of these reactions have enormous potential utility, as small-molecule tools that can tease apart a complex biological system, or as lead compounds that can be developed into the next 'blockbuster' drug. Without asymmetric synthesis, some hugely successful drugs — such as AZT for HIV/AIDS, or lovastatin, which reduces cholesterol — would be extremely difficult to make.

Today, a novice organic chemist is like a child in a sweetshop. There are many 'hot' areas to work on: 'green' chemistry, which seeks ways to eliminate environmentally harmful chemicals from common processes; the burgeoning area of nanotechnology, constructing minuscule components for a future age of molecular devices; or maybe just pushing the boundaries of chemical space, by concentrating on a single element such as boron and seeing what chemistry can be developed. The modern world requires medicines, agrochemicals and advanced materials, and it is chemists who must provide these. Far from being esoteric, organic chemistry serves global needs. ■

Five years on

Immigration restrictions imposed after 11 September 2001 have eased, but improvements must continue.

The attack of 11 September 2001 changed many things, including the way science is done in America. Before the attack, the United States was a focal point for scientists from around the world who went there to study, conduct research and teach. But in its aftermath, a series of restrictive new immigration measures led to lengthy delays for foreign researchers, and many were turned away. In 2004, we warned that these policies were strangling scientific exchange and could have dire consequences for the nation's scientific leadership (see *Nature* **427**, 181; 2004).

As reported on page 6 of this issue, things have improved, although perhaps not as much as many researchers would have liked. The Department of State has boosted staffing levels at its embassies and consulates, and new computer systems are helping to prevent applications from becoming lost during interagency security reviews in Washington. Waiting times for those reviews are down from months to weeks: in Beijing, a student visa that might have taken six months can now be granted in just ten days.

The government has taken other positive steps, abandoning a plan to force foreign scientists to obtain licences to operate lab equipment (see *Nature* **441**, 679; 2006), and revising a rule that would have required foreign-born researchers on defence-funded projects to work in 'segregated areas'. These moves were in response to a vocal

and concerted campaign by a coalition of scientific, university and industry organizations that warned policy-makers that such restrictions would not be in the nation's best interest.

Researchers entering the United States today still find it a hassle, but most feel that it is worth the trouble to come to US laboratories. New statistics are showing a resurgence of foreign students, and researchers are probably following, although numbers reflecting their movement are more difficult to obtain.

This could lead some to believe that the problem is largely solved. But more should be done to ensure free scientific exchange between the United States and other countries. Scientific groups in America should continue to press for better training of embassy staff, and for the hiring of more scientifically aware case-workers. They should work with the State Department and the Department of Homeland Security to try to bring some much-needed transparency to the visa process. It is unlikely that the security checks at the heart of many researchers' delays will become more open, but more could be done to communicate with those entering the United States about the status of their applications. Finally, scientific organizations should support immigration legislation that would make it easier for foreign scientists trained in US universities to remain after finishing their work.

All this should be done with an eye to keeping an open dialogue between scientists and the federal officials overseeing immigration. In the event of another attack on the United States, such lines of communication could help to ameliorate the immigration restrictions that would almost certainly follow. The United States is a linchpin of the global scientific enterprise, and it is in everybody's interest that it remain open for business. ■

RESEARCH HIGHLIGHTS

Best-laid plans

Proc. R. Soc. B doi:10.1098/rspb.2006.3676 (2006)

When your offspring take as long as 17 years to mature, choosing where to lay your eggs is crucial. Louie Yang of the University of California, Davis, studied how cicadas pick their spot.

Cicadas lay their eggs in incisions in tree branches (pictured); the nymphs, once hatched, live underground. To find out what attracts the insects to certain trees, Yang planted 32 red oak saplings in a forest in Virginia. Each tree received different levels of natural light, and was fed with different amounts of nutrients.

Yang observed that insects from the 2004 brood of periodical cicadas chose their trees by light levels, rather than nutrient availability. He speculates that this is because sunlit sites have richer underground environments, dense with roots on which the nymphs can feed.



ROYAL SOC.

GENETICS

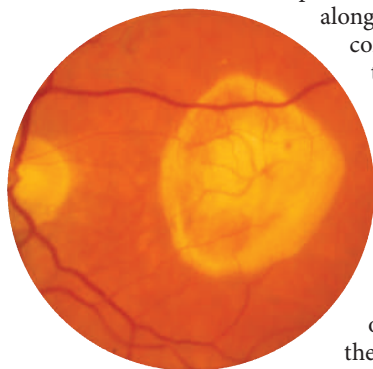
Retinas at risk

Nature Genet. **38**, 1049-1054; 1055-1059 (2006)

Susceptibility to age-related macular degeneration, the most common cause of blindness in the elderly, may be boosted by a number of different changes in one of the genes implicated, report two research groups. The results could be used to screen for those who are at risk.

Earlier studies had linked the chance of developing this disease of the retina (pictured right) to one protein-altering change in an immune-system gene called complement factor H.

Gonçalo Abecasis and Anand Swaroop at the University of Michigan, Ann Arbor, and their colleagues found a further 20 single-letter changes in this gene that are associated with disease risk, but lie outside protein-coding regions. The second team, led by Mark Daly and Johanna Seddon of Harvard Medical School, Boston, confirmed one of these variants in a different sample of people, and also found links between two other genes and the disease.



such as calcium and magnesium.

Uwe Bunz and his co-workers at the Georgia Institute of Technology, Atlanta, designed the molecules to undergo a dramatic colour change on binding to a metal ion. The trick lies in the layout of the molecule's orbitals, between which electrons flit when the molecule fluoresces.

The cruciform shape allows the two important orbitals to be arranged along separate axes. In a conventional fluorophore, the binding of a metal ion can affect both the relevant orbitals; in this criss-cross arrangement it affects only one. This exaggerates the change in the energy gap between the two orbitals, meaning that the shift in the colour of the light emitted is more pronounced than in a conventional fluorophore.

MOLECULAR BIOLOGY

Pumping ions

J. Mol. Biol. doi:10.1016/j.jmb.2006.07.006;

doi:10.1016/j.jmb.2006.07.081 (2006)

Scientists think they may have identified a novel mechanism in a protein that regulates the flow of ions across cell membranes.

The 'chloride pump' CLC-ec1, a membrane protein that helps to control cellular pH, normally exchanges two chloride ions for one proton. However, it can be 'uncoupled' in

such a way that chloride alone flows through.

In two back-to-back papers, Christopher Miller from Brandeis University in Waltham, Massachusetts, and Raimund Dutzler from the University of Zürich, Switzerland, and their colleagues report crystal structures of mutant proteins that are permanently uncoupled.

These findings suggest that protein movement requires chloride to occupy the same active site on the pump that protons use. This is surprising because other exchange transporters studied so far have mutually exclusive binding sites for their ions.

CELL BIOLOGY

Muscle building made easy

J. Cell. Biol. **174**, 677-687 (2006)

Researchers at the University of Virginia in Charlottesville have found that microRNAs can promote the differentiation of some kinds of stem cell into specific cell types.

miRNAs are snippets of RNA that suppress the expression of certain genes by interacting with messenger RNA (mRNA). Anindya Dutta and his group looked at how genes were switched on and off when certain miRNAs were put into muscle stem cells, or myoblasts. They showed that gene regulation by miRNAs encouraged the stem cells to lengthen and stop dividing, as happens when the cells differentiate.

Dutta's team also found hints that some of the miRNAs function by cutting up the target mRNA. Most miRNAs in animals are thought to work by attaching to mRNA and blocking its translation, while leaving the mRNA intact.

P. PARKER/SPL

CHEMISTRY

X marks the spot

J. Am. Chem. Soc. doi:10.1021/ja061112e (2006)

Cross-shaped fluorescent molecules could be turned into powerful sensors for metal ions that play important roles in the cell,

GENE THERAPY

T cells target cancer

Science doi:10.1126/science.1129003 (2006)

In the first successful use of gene therapy to treat cancer, scientists have modified the T cells of cancer patients to recognize melanomas. The result: in 2 of 17 cases, tumours shrank and the patients remained disease-free for more than a year.

T cells recognize their targets via specific receptors. Steven Rosenberg and his colleagues at the National Cancer Institute in Bethesda, Maryland, cloned two T-cell receptors that recognize melanoma tumours. They then extracted T cells from melanoma patients, inserted the gene that would cause them to express the melanoma-targeted receptor, and put the engineered cells back into the patients.

The researchers also cloned two receptors that, *in vitro*, recognize other tumour types, opening the door for future clinical studies.

CHEMISTRY

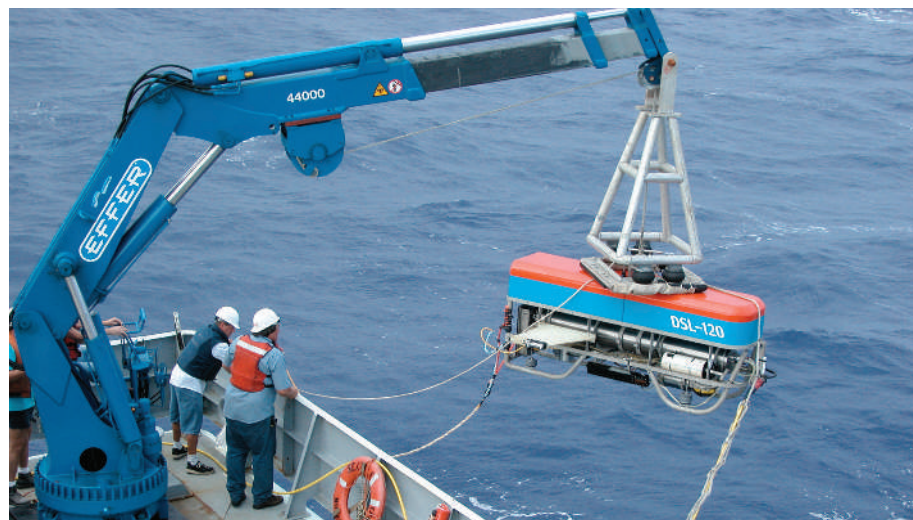
Sweet trick

Science **313**, 1291–1294 (2006)

Some enzymes that attach sugars to molecules during chemical synthesis have been found to have another sweet trick — they can remove sugars too.

Jon Thorson from the University of Wisconsin–Madison and his colleagues show that certain glycosyltransferase enzymes can catalyse reversible reactions in both directions. These reactions can be driven to add, remove or exchange sugar structures as desired.

This surprising discovery reveals glycosyltransferase catalysis to be a versatile tool for the synthetic chemist. To illustrate the point, the researchers report the efficient synthesis of more than 70 variants



W. SAGER/TAMU

of the anticancer compound calicheamicin, each of which has slightly different sugar groups attached.

MATERIALS SCIENCE

Locked up in chains

Nature Mater. doi:10.1038/nmat1726 (2006)

The origin of the bright light produced by blue and green light-emitting diodes is something of a mystery: the nitride semiconductors used to make the diodes contain relatively large numbers of crystal defects, and experience with similar alloys suggests that this should hamper their light-emitting potential.

Now a study suggests that the answer lies with chains and tetrahedra of indium and nitrogen atoms that form in the alloys fortuitously. Shigefusa Chichibu of the University of Tsukuba, Japan, and colleagues show that charge carriers in the alloy, which generate the light emission, are trapped by these chains and thus prevented from interacting with the crystal defects.

GEOPHYSICS

Not-so-quiet zone

Geology **34**, 789–792 (2006)

A putative ‘quiet’ period in Earth’s magnetic history may not have been that quiet after all.

Previous studies suggested that the world’s oldest ocean crust was marked by few of the magnetic-field reversals that pepper modern oceanic crust.

But researchers led by Maurice Tivey of the Woods Hole Oceanographic Institution in Massachusetts argue otherwise. They combined measurements from deep boreholes in the Pacific’s crust with those obtained by towing a magnetometer (pictured above) close to the ocean floor.

The researchers say they have found evidence for reversals in the polarity of Earth’s magnetic field as far back as 170 million years ago, during the Jurassic period. That period seemed quiet, they say, because the polarity changes were of low amplitude and occurred rapidly, making them difficult to detect.

JOURNAL CLUB

Thomas A. Moore
French Atomic Energy
Commission, Saclay, France, and
Arizona State University, USA

A biochemist seeks to better plants’ photosynthetic powers.

Photosynthesis uses sunlight to oxidize water, then recycles the electrons removed in this process to reduce carbon dioxide to energy-rich carbon compounds. If this process could be driven at a rate that matched humans’

fuel consumption, we would have a sustainable energy loop. Photosynthesis in plants is not very power efficient; our hope is that artificial photosynthesis might lend a hand.

To do so, light must be used to produce both powerful oxidizing and reducing species. Photoexcited dyes can perform the trick, provided their oxidation and reduction energy can be captured and directed to do chemical work.

Electron injection into the conduction band — a high-energy level — of titanium dioxide

semiconductor electrodes captures the reducing potential of the excited dye. Unfortunately, things have not been so positive on the oxidizing side — we lack a means to couple the oxidizing potential of a photoexcited dye to water oxidation.

One hope, by analogy with the reducing side, is that the oxidation potential of a photoexcited dye can be captured (through a process known as hole injection) in the lower-energy level of a semiconductor, called the valence band.

Magnus Borgström of Uppsala University, Sweden, and his team recently did this, reporting dye-sensitized hole injection in a porous nickel oxide electrode (M. Borgström *et al.* *J. Phys. Chem. B* **109**, 22928–22934; 2005).

Although their work was geared towards the development of solar cells, I believe that demonstrating the generation of oxidation potential in an electrode holds promise for artificial photosynthesis. The challenge remains to couple the electrode to water oxidation.

SPECIAL REPORT

Safe passage

Lengthy visas delays and persistent security checks have turned foreign scientists away from the United States. Now the country is striving to woo them back. **Geoff Brumfiel and Heidi Ledford report.**

When *Nature* first contacted Olexei Motrunich two-and-a-half years ago, he was beside himself. The 30-year-old Ukrainian physicist had studied and worked in the United States since 1994, and had just taken a postdoctoral position at the University of California, Santa Barbara. But when he travelled home to visit his parents in July 2003, he found himself trapped — his application for a visa to re-enter the United States had disappeared into a mysterious web of post-September 11 security checks (see *Nature* 427, 190; 2004). Today, Motrunich is back in America. His visa was renewed after a six-month delay, but exactly what sparked the background check is still a mystery.

Five years after the terrorist attacks of 11 September 2001, foreign scientists are reporting fewer problems in trying to enter the United States. Waiting times are down and, according to a survey by the Washington-based Council of Graduate Schools, admissions in the sciences are rebounding. Yet anger and unease continue to cloud many researchers' views of America. "It's very hard to overcome the perception that's developed over the past couple of years," says Debra Stewart, the council's president.

US immigration changed suddenly and dramatically after the 2001 attacks. Congress passed legislation requiring face-to-face interviews with every visa applicant, leading to lengthy delays — even in European countries where the visa process was traditionally smooth. Scientists had longer waits than most because their fields of study often appeared on the government's Technology Alert List. Many researchers' applications were sent to Washington for background checks that could take months, according to Barry Toiv, director

of public affairs for the Association of American Universities, a Washington-based group that represents university interests. "We went through a very difficult period following 11 September," he says.

The situation has improved greatly since then, according to Tony Edson, deputy assistant secretary of state for visa services at the US Department of State. Embassies and consulates have increased staffing, and new computer systems have been installed that allow applications to pass electronically between embassies and Washington agencies. To reduce the number of security checks, consular officers are receiving some additional training in handling scientific cases, and they are being encouraged to consult science attachés when appropriate. "We've added extra people, we're investing in infrastructure," Edson says. "That has improved the situation."

Waiting game

The average wait still varies widely from country to country, but state-department statistics on the length of the Washington-based security checks that many scientists encounter show dramatic improvement. The average wait has dropped from 2.5 months in 2003 to two weeks by last December.

But those statistics do not reflect the trouble individual scientists can have when entering the United States (see 'The cold shoulder'). When Goverdhan Mehta, a chemist and former director of the Indian Institute of Science in Bangalore, applied for a visa to become a visiting professor at the University of Florida at Gainesville in February, embassy officials delayed his application and badgered him



about how his research might relate to chemical weapons (see *Nature* 439, 901; 2006).

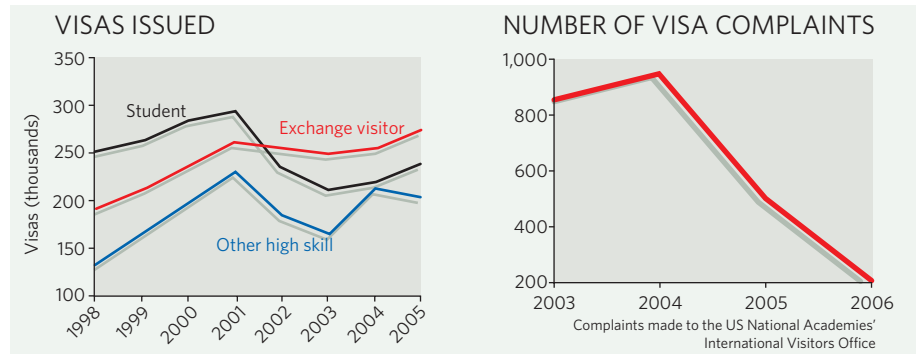
Mehta was so incensed that he declined the visa when it was eventually offered to him. The University of Florida has renewed its invitation, but Mehta remains ambivalent. "Certainly I am not going to subject myself to the same process," he says.

Stewart says that such stories are still damaging the United States' image in the international scientific community. "Every time there is a high-profile case like that it's five steps backwards," she says.

To try to improve the views of foreign students and scholars, many US universities are now reaching out to the international community, according to Marlene Johnson, executive director of the Association for International Educators. International student offices have boosted staffing and many universities have opened recruiting offices in countries such as China. "I don't know of a single major research university that was recruiting before 9/11," Johnson says.

The increase in recruitment and drop in waiting times seems to have had a positive effect (see graph). According to a survey by the Council of Graduate Schools, released last month, graduate admissions in the 2005–06 school year were up sharply. Places offered to students from India and China, the largest suppliers of science and engineering students, rose 28% and 20%, respectively. That may have

SOURCES: NSF & NAS





US VISAS IN FOCUS
Nature's collection of news
on immigration controls.
www.nature.com/news/infocus

R. TALAIE/CORBIS
NG HAN GUAN/AP



US security measures have left many foreign scientists tied up in red tape.

contributed to a less impressive, but still positive, 1% rise in admissions to PhD programmes in the life sciences and a 5% rise in physical sciences. Final enrolment numbers out later this year are expected also to show a positive trend, says Stewart. "We're seeing a turnaround."

Not reflected in the numbers are those, such as Mehta, who have given up on coming to the United States. "The visa trouble was definitely one reason why I went back to Germany," says Stefan Gilb, a chemistry professor at the Technical University of Munich, whose visa renewal was delayed during his time at the University of California, Berkeley. And Reza Mansouri, an Iranian physicist, was pulled out of line by security in May at a US consulate in Montreal, Canada. After several hours of delays he decided to go home rather than be hassled further. "I told them 'I don't need a US visa. Bye-bye,'" he says. Mansouri did eventually get his visa, however, and is due to arrive in San Francisco this week.

Most education experts suspect that, despite delays, scientists and graduate students will ultimately come back to the United States. "We have 4,000 institutions and the capacity to absorb a whole lot of international students," says Peggy Blumenthal, a vice-president at the Institute of International Education, a group based in New York that tracks the flow of foreign students and scholars. Well-stocked

labs with strong funding opportunities mean that "particularly in fields such as science and engineering, we will continue to attract the best and the brightest", she adds.

But America's hegemony is not ensured. In Australia, overseas enrolment has more than doubled from 2000 to 2004, and, according to a recent government report, today roughly a quarter of its one million undergraduate and graduate students come from outside the country. Significant rises have been reported in Europe, and China and India are attempting to expand their domestic higher-education sectors, especially in science and engineering. These trends were already apparent in the late 1990s, and in a way, says Stewart, the restrictions since 2001 may have helped the United States come to terms with the increasingly competitive global market. "We still have the best doctoral programmes in the world," she says. "But it's not our God-given right."

For his part, Motrunich has decided to stay, at least for now. He's leaving the University of California, Santa Barbara, to begin a new job at the California Institute of Technology in Pasadena this autumn. Whatever his problems with the US immigration system, his colleagues have never made him feel unwelcome, he says. "In the scientific community, the fact that one is from a foreign country plays no role."

See Editorial, page 2.

The cold shoulder

Stefan Gilb

For chemist Stefan Gilb, the changes to US visa policy have been too little, too late. Three years ago, Gilb flew home to Germany to renew his visa before beginning a postdoc position at the University of California, Berkeley. Gilb expected the process to take two weeks, but was stuck in Germany for three months. Gilb thinks the consulate in Frankfurt sent his application for added security checks because his work involved lasers. "I told them what research I was doing and they put me on the list," he says. "Nobody would tell me how long it was going to take."

A year later, when Gilb had to renew his visa again, the process took just a week. But he has since returned to Germany, to the Technical University of Munich, a decision he says was strongly influenced by his visa troubles.

Olexei Motrunich

Ukrainian physicist Olexei Motrunich was also stuck in his home country for months awaiting a visa. In 2004, he waited six months to begin his postdoc at the University of California, Santa Barbara. Motrunich wasn't given an explanation either, but suspects it was related to his research at the time, which included work with satellites and neutron interferometry.

The visa difficulties didn't deter him from deciding to stay in the United States, however, and he is about to start a new job as an assistant professor at the California Institute of Technology in Pasadena.

Goverdhan Mehta

Indian chemist Goverdhan Mehta is a former director of the Indian Institute of Science in Bangalore, current director of the International Council for Science, and has been a science adviser to India's prime minister.

Mehta has a number of collaborators in the United States, and has travelled there on a visitor's visa. But when he applied for a visa to become a visiting professor at the University of Florida at Gainesville, embassy officials delayed his application and questioned the nature of his research. Mehta describes the experience as "humiliating".

The debacle created a diplomatic wrinkle just before President Bush's first visit to India. Eventually, the US embassy in India issued a statement of regret, and when Mehta's visa was approved, US ambassador David Mulford informed him personally. Mehta declined the visa. The University of Florida has renewed its invitation, but Mehta has not yet decided whether or not he will accept. "The experience was too rattling," he says. "It will take time to heal."

H.L.

Mix and match: the hunt for what makes us human

Genome studies are turning up exciting hints about what makes human brains unique. But scientists must now tackle a difficult question: how can they tell which leads point in the right direction and which are red herrings?

The new studies rely on genome sequence data from humans' closest relatives: other primates. A draft of the chimpanzee genome was released last year¹ and a draft sequence of the macaque genome is also publicly available. The data are enabling scientists to compare chimp, human and monkey DNA to search out signals that seem unique to people.

This approach is yielding pointers to the mechanisms that may have driven human evolution. Last week, researchers led by James Sikela of the University of Colorado Health Sciences Center in Aurora, announced the latest². They looked for genes repeated more often in human DNA than in chimp genomes. They found that one gene contains a protein-coding domain that is repeated 212 times in people, compared with 37 repeats in chimps, 30 in macaques and one in mice and rats. And, significantly, this protein domain is present in human brain cells. It is also present in other tissues, but Sikela's team is excited because the domain is

found in the neocortex, which is much larger in humans than in other primates.

"The fact that a huge explosion of this domain occurred in the primate lineages, and the fact that it is found in the brain, seems to make it a good candidate for cognitive function," says Sikela.

Investigating such 'copy-number polymorphisms' — gene repeats, deletions and other large-scale rearrangements — is one of the hottest areas of genomic research, says geneticist Evan Eichler of the University of Washington in Seattle. These sorts of structural variants have only recently been discovered, and it's possible that they could be a major driver of human evolution³. "One take-home message here is that duplications rock, and structural variation is very important," Eichler says.

A broader question raised by Sikela's work and other recent studies is how to find out what, exactly, the newly discovered genes are doing. The easiest way to pin down an unknown gene's function — mutating it in an individual — isn't ethically acceptable in chimps or, of course, people.

One solution, says Sikela would be to mutate the relevant genes in mice, to see what happens

"One take-home message here is that duplications rock."



to the rodents' brains. In a 2002 study, researchers at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, found that they could enlarge mouse brains by inducing fetal mice to over-express the regulatory chemical β -catenin⁴. That suggested the chemical might regulate pathways that control human brain size.

Researchers can also scan tissue banks to find out when in human development genes are expressed. In a paper published on 16 August⁵,

Muslim council phases in lunar calendar

The Muslim community in the United States has taken a small but significant step towards resolving one of the oldest disputes between Islam and science — the creation of a unified lunar calendar.

Islam, in common with Judaism, uses lunar dates worked out in accordance with the phases of the Moon. But despite major advances in lunar astronomy over the past few centuries, Muslims have never agreed a single lunar calendar. This may finally be about to change. At a meeting last month, the Fiqh Council of North America, a body of American Islamic religious scholars, agreed that religious festivals in the United States such as Eid, the end

of the fasting month of Ramadan, will now be fixed according to a predetermined calendar.

In a statement on 28 August, Muzammil Siddiqi, chairman of the Fiqh Council, said: "Muslims will be able to plan their activities in advance, take time off from work or school. It will reduce a lot of chaos, hardship and confusion."

The move was welcomed in Britain by Zafar Iqbal who chairs the Islamic Calendar Committee of the Muslim Council of Britain, a coalition of



The timing of Islamic festivals in the United States will no longer rely on direct sightings of the Moon.

some 400 mosques and community groups. Iqbal believes that the US decision will strengthen calls for Britain's estimated 1.6 million

Muslims to adopt a unified calendar. "This is a very courageous decision," he says.

The US decision has its critics among both scientists and theologians, including members of the Fiqh Council itself. One member says he doesn't think the move will have universal appeal — especially among those who believe that it is a religious requirement for the beginning of each lunar month to be confirmed by a naked-eye sighting of the new Moon.

This need to see the new Moon dates from the time of the Prophet Mohammed in the seventh century. And there is a reluctance to switch from what people believe

D. SANSONI/IMPACT PHOTOS

**CHIMP SPECIAL**

Discover the similarities and differences between chimps and mankind.

www.nature.com/news/specials/chimpgenome



mental retardation. One patient had a deletion in the same region of the genome where Sickle's group found its new gene. "It doesn't prove anything, but it's interesting," Eichler says.

Finding the function of highly repetitive genes is especially challenging, he adds, because such genes are tricky to analyse correctly, and genome assemblers often miss them. So there are large gaps in our understanding of where duplications occur. Eichler is trying to remedy this with a project to sequence all the structural variations in a small group of people.

Other investigators say that it will be hard to make better sense of the new genome studies without more studies on apes. Because chimps are difficult to work with, and scarce, researchers don't really know much about how the structure and function of chimp brains compare with those of human brains, says Todd Preuss, a neuroscientist at the Yerkes National Primate Research Center in Atlanta, Georgia. That makes it hard to decipher the meaning of genetic disparities in the two species' brains, and that must change, he says. "We're all dressed up with no place to go — we've got all this wonderful information and we can't do much with it. The way to get around this is to study apes, and to start looking at the differences between human and chimp brain organization." ■

Erika Check

1. The Chimpanzee Sequencing and Analysis Consortium *Nature* **437**, 69–87 (2005).
2. Popesco, M. C. *et al. Science* **313**, 1304–1307 (2006).
3. Check, E. *Nature* **437**, 1084–1086 (2005).
4. Chenn, A. & Walsh, C. A. *Science* **297**, 365–369 (2002).
5. Pollard, K. S. *et al. Nature* doi:10.1038/nature05113 (2006).
6. Sharp, A. J. *et al. Nature Genet.* **38**, 1038–1042 (2006).

Researchers are comparing human and chimp genomes to find out what makes our brains so different.

a team led by David Haussler at the University of California, Santa Cruz, did just that. By comparing the human and chimp genomes, they found one gene that has evolved rapidly during the transition from chimps to people. They looked at embryonic tissue collected by collaborators in Belgium, and found that this gene is expressed at 7–19 weeks of human

development, when neurons are known to be forming and moving throughout the brain.

Another way to work out what brain genes do is to look for effects on human health, and to study patients with brain dysfunctions. In a study published on 13 August in *Nature Genetics*⁶, Eichler's lab looked at duplications in the genomes of 290 patients with unexplained

Mohammed required, regardless of whether a scientific approach might be more convenient — in fact doing so is seen by many Islamic scholars as coming close to committing a sin.

Usama Hasan, a lecturer in computer science at Middlesex University, UK, is an Islamic scholar and has collaborated with the Royal Observatory in London. He says that the council should have used a more accurate method to compile the calendar. The council opted for a method based on computing the position of the Moon relative to Earth and the Sun. But Hasan says that the technology now exists to predict whether the Moon will be visible in a certain place at a certain time, which comes closer to the requirements of Islamic tradition. "Seeing the Moon with the naked eye is a means to an

end; it should not be seen as an end in itself," he says.

But the naked eye remains the method of choice throughout the Islamic world, and is one of the main reasons a unified calendar has so far proved elusive. According to this tradition, a new lunar month begins the morning after a sighting of the thin crescent Moon — some 24 hours after the birth of a new Moon.

That's difficult for countries in the Northern Hemisphere, because thick clouds often hide the crescent Moon. And in countries at higher latitudes, such as in Scandinavia, the crescent is invisible to the naked eye. When this happens, mosques

in northern Europe and North America tend to follow decisions in other countries. Saudi Arabia and Pakistan are favourites, but even they often celebrate Islamic festivals on different days.

Hasan is confident that the US decision will prompt a rethink in Saudi Arabia, where Moon-sighting causes frequent controversy. Saudi Arabia does have an Islamic calendar based on lunar tables, known as

the *Umm al-Qura* calendar. This was developed by the King Abdulaziz City for Science and Technology, the country's science ministry, and is used for all non-religious purposes. But human sightings are still used

"Seeing the Moon with the naked eye is a means to an end; it should not be seen as an end in itself."

to determine dates for religious festivals. Moreover, anyone can claim to have seen the Moon — the authorities are duty-bound to listen, and sometimes accept news of a sighting even if this does not agree with the dates in the lunar calendar.

This happened last year, when the Saudi government decided, with just ten days to go, to revise the date for the annual Hajj pilgrimage because of a claimed Moon sighting — even though astronomical calculations showed such a sighting to be impossible. The sudden change led to chaos in the organization of the three-day event, which attracts 2 million people from more than 150 countries.

Siddiqi, too, is hopeful that the use of lunar calendars will catch on, if only to rule out erroneous sightings. ■ Ehsan Masood

When science and theology meet

Religion is religion, science is science, and good fences make good neighbours. That seems likely to be the thrust of an expected clarification by the Roman Catholic Church of its position on biological evolution, according to a prominent biologist who spoke at a retreat on the topic held last weekend by Pope Benedict XVI.

This outcome would be a blow to supporters of intelligent design who had hoped that factions in the Vatican sympathetic to their ideas might prevail in an internal feud over evolution that has opened since the death of Pope John Paul II in April last year. That could have given a global axis to what has been largely a US cultural phenomenon.

John Paul II perhaps did more than any pontiff to reconcile faith and science, declaring in 1980 that there was no contradiction between the two. He subsequently described darwinian evolution as more than a hypothesis, and in 1992 officially rehabilitated Galileo Galilei, admitting the Church's fault. Benedict XVI, however, has long been more ambiguous on the topic. In his first mass as Pope, he stated that "we are not some casual and meaningless product of evolution".

Last year, the prominent Cardinal Christoph Schönborn wrote an article in *The New York Times* expressing doubts about darwinism. He also seemed to advocate intelligent design, a movement whose advocates seek to give creationism scientific credibility by arguing that the complexity of science can be explained

only by divine intervention in the process of evolution.

Although Schönborn has since backpedalled on the way his ideas were expressed in the article, the piece fuelled rumours that Rome, which currently endorses darwinian evolution and rejects any literal interpretation



Creative thought: Pope Benedict XVI is wrestling with how to square evolution with religion.

French confusion leaves Iranians locked out of meeting

PARIS

Global politics has soured the run-up to this week's TNT2006 Trends in Nanotechnology conference in France. Organizers initially banned some 15 Iranian researchers from attending, citing political concerns over Iran's nuclear programme.

The decision has since been revoked and two Iranians will now attend the conference on 4-8 September in Grenoble. The remainder have withdrawn, according to the organizers, because of problems such as visa delays. But the incident has

raised concerns that the current political situation, in which Iran faces international isolation over its refusal to stop its uranium enrichment programme, may cause scientists from the country to be ostracized.

Those involved in the conference have denied any discrimination against Iranian scientists, and say the incident is the result of a series of misunderstandings about the security-clearance rules. The congress centre in which the conference is being held is inside Minattec, a sprawling research and industrial complex that belongs

to the French Atomic Energy Commission (CEA).

When Antonio Correia, president of conference organizers the PHANTOMS Foundation in Madrid, began receiving submissions from Iran he contacted Didier Molko, CEA director of Minattec, to check whether this raised any security concerns. Molko told him that "CEA rules say that Iranian people can't go in". As a result, Correia e-mailed the Iranian applicants on 9 June, apologizing that "we must withdraw your contribution and participation", a move he says met an angry response. "I understand

that people were angry," says Correia. "Science should have nothing to do with politics."

Molko now says his initial decision was a personal mistake, and not CEA policy. He was being cautious because of the political situation, he says, but after checking with his superiors in Paris, he revoked his earlier decision. "It was a delay of a week or two at most."

Pascal Newton, a spokesman for CEA headquarters in Paris, says that the commission's security services check out all visitors to CEA buildings on a case-by-case

of biblical creation, might instead endorse intelligent design.

Speculation increased last month when Schönborn announced that creation and evolution would be the topic of the Pope's annual retreat, to be held in Castel Gandolfo near Rome. As a former cardinal and German theology professor, Joseph Ratzinger has long held this informal gathering, where he discusses a hot topic with his past students. But the retreat has taken on new significance since Ratzinger became Pope.

Bruce Chapman, president of the Discovery Institute in Seattle, Washington, which represents the intelligent-design movement, welcomed the choice of topic in a press release published on the institute's website last week. In it he wrote: "What's good about all this is that they're taking up the issue seriously...it will be conducted at a very high level and I think it should give cheer to people who are critics of Darwinism."

Schönborn was one of four invited speakers at the meeting, which also included Robert Spaemann, a conservative German philosopher, and Paul Erbrich, a Jesuit priest who questions the random nature of evolution. The fourth speaker, the only working scientist present, was Peter Schuster, a molecular biologist and president of the Austrian Academy of Sciences.

In a break with tradition, the proceedings of the meeting will be published later this year, says Schuster, with a preface written by the Pope. The message will be to promote dialogue between faith and reason, Schuster says. Given the power struggles within the Church,

however, the precise outcome of the overall debate is impossible to predict, he says: "We have to wait."

But discussions at the meeting suggest that the Church will probably affirm a form of theistic evolution, which posits the general principle that biological evolution is valid, although set

in motion by God. At the same time, it seems likely to reject the fundamental intelligent-design principle that God was a watchmaker, intervening in the details. "Intelligent design as an intervention of God during evolution will not be an outcome," predicts Schuster. "I got the impression that there was general agreement that

evolutionary biology is a undeniable science and not a hypothesis."

The Pope in particular "immediately accepted that theology is not going to interfere with science", Schuster adds. "He is not a scientist, but I was surprised by the sharpness of his intellect. He wanted to be informed; he was very interested in science."

But although there is likely to be agreement that the Church should not attempt to question the basic theory of evolution, those at the meeting were also adamant that science should not stray into theology. Schuster says there was a wide desire to address the Church's concern that darwinism is being extrapolated as a wider metaphysical stance. This argues that we are random products of evolution, and that there is therefore no need for God. "I agree that, as in science, everyone should stay within their realms of competence," says Schuster. ■

Declan Butler

basis. "We get a simple, yes/no; we never know why." Particular nationalities aren't officially banned, he says, although he adds: "In the current political context, I can imagine Iran might be more sensitive."

The United Nations Security Council set a deadline of last week for Iran to suspend uranium enrichment activities, which it suspects are part of a programme to develop nuclear weapons. Iran ignored the deadline and, failing last-ditch

negotiations this week, the council is set to impose economic sanctions on the country.

International isolation in similar situations has in the past spared scientific cooperation. Indeed, during the cold war, contact between scientists was one of the few channels open between East and West. But Reza Mansouri, a well-known Iranian physicist based at the Sharif University of Technology in Tehran, complains that his

country's scientists are already falling victim to discrimination, such as visa delays. He also worries about what he sees as widespread ostracization by fellow scientists, and he deplores the incident in France.

But Mansouri says he believes it is a temporary problem induced by the tense international climate. "In the long term, the best scientists will continue to collaborate with our scientists, bringing more rationality and understanding into the dialogue, and ultimately more peaceful attitudes." ■

Declan Butler

ON THE RECORD

"They are wanted to be kissed. This is the reason why the two young ladies are invited."

Igor Tkatchenko, vice-chairman of the French Chemical Society, bringing on stage two women chemists who got honourable mentions at the European Young Chemist Awards in Budapest.



"I could argue the case with hand puppets and win."

Attorney David Bookbinder on the argument that the US Environmental Protection Agency should restrict emissions of greenhouse gases.

"ETs...could also be creatures of God."

Father Jose Funes, new director of the Vatican Observatory, ponders the nature of intelligent life elsewhere in the Universe.

NUMBER CRUNCH

15% of HIV-positive people in the world live in India.

0.9% of Indian adults have HIV.

64% of Indian lawmakers believe HIV can be passed on by sharing clothes with an infected person.

SCORECARD



Iraq archaeology

He stayed longer than many Iraqi academics, but Donny George, the country's most prominent archaeologist, has finally given up. George was known for his efforts to recover antiquities looted after the war in 2003. But he has now resigned and fled to Syria, citing security problems and lack of funding.



Blue skies research

Scientists in Britain will need to fight a little harder to justify their funding. UK research councils will now include wealth creation experts on all funding panels, to help judge which projects deserve to receive the money.

Sources: San Francisco Chronicle, Indian Catholic, Indian Association of Parliamentarians, UNAIDS

'Ethical' stem-cell paper under attack

"What we have done, for the first time, is to actually create human embryonic stem cells without destroying the embryo itself."

So stated Robert Lanza on the *Nature* podcast on 23 August, when his group's research paper — whose summary implied the same achievement — was published online (I. Klimanskaya *et al. Nature* doi:10.1038/nature05142; 2006). The paper attracted worldwide media excitement, particularly in Germany, where the law prevents scientists from working on newly derived human embryonic stem-cell lines. Shares in Lanza's company, Advanced Cell Technologies (ACT) in Worcester, Massachusetts, leapt five-fold in price in just 10 hours.

But this was quickly followed by a backlash as it became clear that all 16 of the embryos used in the research had been destroyed. German newspapers, for example, accused Lanza, as well as *Nature*, of hyping the results.

"So many things about the paper and how it was presented are unclear," complains stem-cell scientist Hans Schöler, a director of the Max Planck Institute for Molecular Biomedicine in Münster, Germany. Schöler says he is disappointed that such confusion should cloud the field so soon after the Woo Suk Hwang scandal, in which the South Korean scientist falsely claimed to have created embryonic stem-cell lines from cloned human embryos.

No one is suggesting that Lanza's paper is in any way scientifically incorrect, but many in the field have objected, like Schöler, to its overall packaging. "Perhaps 'word-smithy' is the right term," comments George Daley of the Harvard Stem Cell Institute.

As well as the comments on the podcast, the paper's abstract also implied that the ACT scientists had created stem-cell lines without destroying embryos, saying: "The ability to create new stem-cell lines and therapies without destroying embryos would address the ethical concerns of many." *Nature*'s press release also stated that only one cell had been taken from each embryo — a misconception repeated in subsequent news stories.

The paper itself, however, shows only proof of principle that a human embryonic stem-cell line can be created from a single cell, or blastomere, from a very early embryo comprising 8–10 cells. None of the embryos used survived. The *Nature* paper's details and Supplementary Information made clear that all the embryos were broken up, stating that 91 cells were used from 16 embryos. Only two stem-cell lines were created,



R. FRIEDMAN/CORBIS

Robert Lanza believes his human embryonic stem-cell research will resolve ethical concerns.

so the efficiency of the process — barely 2% — is much lower than it appears to be at first glance.

Some feel this caveat was hidden too deeply. Peter Andrews of the University of Sheffield, UK, co-leader of ESTOOLS, the European Union's programme to develop human embryonic stem-cell biology, says that researchers as well as journalists were fooled: "Many scientists at the kick-off meeting of ESTOOLS a couple of

days later had quickly read the manuscript and got the impression that Lanza really had extracted just a single cell, and that the embryos survived."

Within minutes of the paper going live, *Nature*'s press office corrected its press release to say that Lanza's experiments had destroyed some of the embryos. Two days later, a second note made clear that all the embryos had been destroyed. "*Nature* takes responsibility for the problems with the press release, for which I apologize," says Philip Campbell, editor-in-chief of *Nature*. He adds that a clarification has been posted alongside *Nature*'s podcast, and that the paper will also be clarified.

Lanza says he never intended to say more than that he had proved a principle, and that he is surprised by the reaction to the paper. He adds that the established procedure of pre-implantation genetic diagnosis (PGD), in which a single cell is extracted for genetic analysis, has already shown that embryos from which a blastomere has been removed survive. "We knew that, so we took multiple cells from each embryo so as not to be wasteful," he says.

The *Nature* paper says Lanza's method will "allow the generation of matched tissue for children and siblings born from transferred PGD embryos", and he now plans to take the work into the clinic. He hopes the methodology will provide a source of stem cells compatible with US law, which prohibits public funding of research into new human embryonic cell lines derived at the expense of embryos.

Many stem-cell researchers describe the science in Lanza's paper as interesting, if preliminary. "Anything that you learn about human stem-cell biology, is going to be helpful, whatever way you want to derive new cell lines," comments Daley. But it is still not clear whether applying this approach to PGD embryos would resolve ethical concerns (see *Nature* 442, 858; 2006). Norio Nakatsuji of Kyoto University, who derived Japan's only human embryonic stem-cell lines, points out that the process poses a small risk to the baby. "We don't know how this would be legally interpreted," adds James Battey, chair of the US National Institutes of Health (NIH) Stem Cell Task Force. "The NIH would need a test case."

Scientists in countries where cell lines may be legally derived from embryos discarded from *in vitro* fertilization procedures, such as Britain, Sweden and Japan, also object to the idea of carrying out research just to get round political restrictions. "It is more ethical to work on embryos that are going to be destroyed anyway," says Andrews. "If you do science for the purpose of getting around the law it makes things worse; it makes science look dishonest." ■

Alison Abbott

NIH proposes database for disease-linked gene data

The US National Institutes of Health (NIH) wants to introduce a policy to control the way data from gene-association studies are shared and used by researchers.

Around 20 such studies, which attempt to find the genes involved in health and disease, are under way or in the planning stages by scientists funded by different branches of the NIH. The agency is proposing a central data repository that would broaden access to information and allow it to be translated more quickly into clinical treatments.

Under the proposal, researchers would have to place their data in the repository promptly, removing information that might reveal the identity of their study subjects. They would have a grace period of nine months to be the only ones to publish on their data, and information in the repository could not be patented until products were developed from it.

The NIH is asking the public for comments until 31 October.

Dinosaur count shows how few have been discovered

More than two-thirds of all dinosaur genera are still waiting to be discovered, according to the latest statistical census of fossils.

Just under 530 genera have been scientifically described. In principle, more than 1,300 additional genera could still be unearthed, say researchers who have studied the abundance of genera that have been identified by a single fossil, as well as of genera for which many fossils exist (S. C. Wang and P. Dodson *Proc. Natl Acad. Sci. USA* 103, 13601–13605; 2006).

"It's a safe bet that a child born today could expect a very fruitful career in dinosaur palaeontology," says co-author Peter Dodson, a palaeontologist at the University of Pennsylvania in Philadelphia. "Unfortunately, there is a finite limit



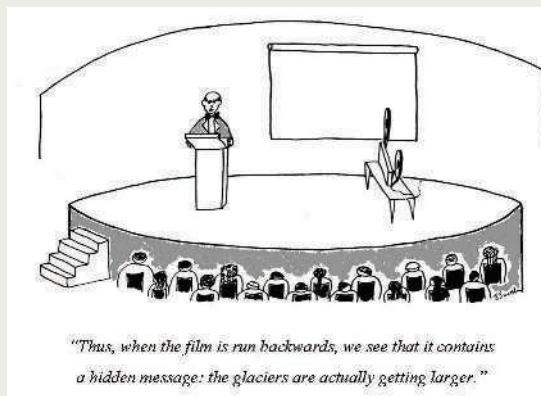
Fossil wealth: palaeontologists can look forward to discovering another 1,300 genera of dinosaur.

Cartoons highlight anger over political interference

US researchers have a new way to fight political interference in science: humour.

The Union of Concerned Scientists, an advocacy group based in Cambridge, Massachusetts, has launched a competition for the best cartoons dealing with the issue. Twelve finalists have been chosen. Many poke fun at the Bush administration's stance on global warming (see right), whereas others deal with issues of research integrity at agencies such as the Environmental Protection Agency and the National Institutes of Health.

The winner will be chosen by popular vote. Anyone can vote for their favourite cartoon at www.ucsusa.org/scientific_integrity/science_idol until the end of the month.



to what can be discovered, so our estimates show that the child's grandchildren won't be so fortunate."

Ion-powered craft makes grand finale on the Moon

It's not often that space scientists celebrate a crash-landing, but they did in the early hours of 3 September, as mission controllers at the European Space Agency (ESA) deliberately smashed the SMART-1 spacecraft into the Moon. The impact, at two kilometres per second, sent out a bright plume of dust and debris that was captured by the Canada–France–Hawaii Telescope on Mauna Kea, Hawaii.

It was a fitting end for the three-year mission, which spent half of that time travelling a looping set of lunar orbits, covering some 100 million kilometres on just 60 litres of fuel. The craft used a solar-

powered ion drive instead of chemical propulsion.

The impact created "a more intense flash than expected", says chief mission scientist Bernard Foing, who watched it from ESA's facility in Darmstadt, Germany. Planetary scientists will now visually analyse the lunar soil ploughed up by the impact to determine its composition.

► www.nature.com/news/specials/moon/index.html

UK outlaws DNA tests performed without consent

DNA theft is illegal in Britain now that the 2004 Human Tissue Act finally became law last week. Full consent must be obtained from a person before any form of test can be carried out on their DNA. The restrictions, however, do not apply to the police.

The act was introduced primarily to regulate organ donations. Researchers will now require an official licence from the Human Tissue Authority to store tissues or cells for genetic testing. Existing samples, taken before 1 September, are unaffected by the act. The new legislation on DNA has particular implications for paternity testing, preventing secret tests to confirm the identity of a child's father.

"It is no longer the case that people can surreptitiously pick up a beer glass that someone has just been using and take a DNA sample from it," says Adrian McNeil, chief executive of the authority.

Californian institutes lured to Florida by cash offers

The Florida gold rush has begun. A US\$310-million package has induced the Burnham Institute for Medical Research in La Jolla, California, to set up a research centre in Orlando, Florida.

Prizewinning mentors

The 2006 winners of the UK Nature/NESTA Awards for Creative Mentoring in Science were announced earlier this week.

The mid-career prize was awarded to Steve Watson, professor at the Department of Cardiovascular Medicine at the University of Birmingham.

There were two joint winners of the award for lifetime achievement: Andrew McMichael, director of the Wetherall Institute of Molecular Medicine, University of Oxford, and Godfrey Hewitt, emeritus professor at the University of East Anglia.

Details of the scheme, co-sponsored by Nature and the UK National Endowment for Science, Technology and the Arts (NESTA), can be seen at www.nature.com/nature/nestaawards/index.html. Details of the achievements of the winners will appear in a forthcoming issue of Nature.

Under the agreement, the state of Florida and local government will provide the institute with buildings, laboratories and operating expenses — even golf-club memberships — for years, as part of a drive to attract high-tech facilities to the state (see *Nature* 442, 729; 2006).

In addition, the Torrey Pines Institute for Molecular Studies, also of La Jolla, has signed a non-binding letter of intent to accept some \$90 million to open labs in Port St Lucie, Florida. But it is still bargaining over where precisely to locate, apparently in a bid to sweeten the deal.

Research submarine sunk by the rising cost of war

The high cost of the war in Iraq has led the US Navy to decommission its last old-style research submarine, shortly after spending \$60 million to refurbish it.

The 38-year-old USS *Dolphin* — which holds the record for the deepest dive by

companies in the state will have to abide by the new requirements, which aim to cut carbon dioxide emissions to 1990 levels by 2020.

Elsewhere, the World Bank has agreed to pay two Chinese chemical firms more than \$1 billion to cut their greenhouse-gas emissions by the equivalent of 19 million tonnes of CO₂ a year. The deal is part of the Clean Development Mechanism of the Kyoto Protocol on climate change, which allows richer countries to meet some of their obligations to reduce greenhouse gases by funding efforts in developing nations.

The agreement will target emissions of HFC-23, a gas formed in making refrigerants that has a global-warming potential 11,700 times that of CO₂.

European research council names its first head

The scientific council of the European Research Council (ERC) has appointed molecular biologist Ernst-Ludwig Winnacker as its first secretary-general. The ERC, which will distribute grants across the continent on the basis of excellence, will be formally established as an agency of the European Commission in January 2007.

Winnacker comes with relevant experience: in December, he will complete a nine-year mandate as president of Germany's main granting agency, the DFG. A leading figure in the decade-long battle to create a European-level agency, he has argued that the ERC should be independent of the commission, to avoid any political influence.

The commission has so far kept its promise not to interfere in the ERC's early procedures, Winnacker says. "But all research agencies have to be watchful of keeping politics at arm's length," he adds.



Down and out: the USS *Dolphin* leaves the US Navy this month after 38 years of service.

a submarine at 1,000 metres — will leave service on 22 September. In 2002, flooding and an onboard fire forced extensive repairs to the vessel (see *Nature* 417, 478; 2002).

The Navy is leasing a Swedish submarine, *Gotland*, for research as an interim measure. Other Navy vessels will also conduct projects previously undertaken by the diesel-electric *Dolphin*; but officials say these will cost more because nuclear subs are more expensive to operate.

Emissions under fire from China and California

Last week saw two steps forward in the battle against climate change, on opposite sites of the planet.

In California — a long-time leader on US pollution controls — state legislators and governor Arnold Schwarzenegger announced a deal that requires industries to cap their greenhouse-gas emissions. All

Corrections

In the News Feature "Caught in time" (*Nature* 442, 736–738; 2006), the graphs showing survival rates for breast and colorectal cancer have incorrect x-axes. The orders of the labels should be reversed, reading 'Distant, Regional/distant, Local', not the other way round.

In the News story "Pluto: the backlash begins" (*Nature* 442, 965–966; 2006), we characterized Mark Sykes of the Lunar and Planetary Science Institute as believing that a better definition of a 'planet' than that of the International Astronomical Union would be hard to come by. This is not his position. He believes that an improved definition would be easily obtained — but that agreement on it within the community will require a more open, inclusive process.

BUSINESS

The farmyard drug store

Pharmaceuticals made in genetically modified animals have been poised to take off for years. **Heidi Ledford** investigates the reality.

The birth of Herman the transgenic bull in 1990 was supposed to herald the start of a new era in pharmaceuticals. As a bull, Herman failed spectacularly in his main job: to produce an iron-rich human protein called lactoferrin in milk. Even his offspring produced disappointingly low levels of the protein.

But hopes have been renewed by Sweetheart, a transgenic goat who lives in Massachusetts, on a farm run by Framingham-based GTC Biotherapeutics. On 2 August, the anti-clotting agent in Sweetheart's milk, ATryn, became the first drug produced by a transgenic animal to be approved by the European Medicines Agency (EMA). Two weeks later, Pharming Group — a descendant of the company that created Herman — announced that the EMA had accepted its application to market an anti-inflammatory drug produced in transgenic rabbits.

The farmyard may seem alive with drug possibilities these days, but the field is still struggling. Sweetheart's success is the first after a series of failures, and some analysts argue that major industry players have lost interest altogether in the pharmaceutical promise of transgenic animals.

"The industry has been holding its breath for over a decade for a product to be approved by the Food and Drug Administration or EMA," says Willard Eyestone, professor at the Virginia-Maryland College of Veterinary Medicine in Blacksburg, Virginia.

Enthusiasm was far higher in the early 1990s, when drug companies first started to focus on biopharmaceuticals — products made from large biological molecules such as proteins and antibodies. Proteins are typically produced in giant vats of cultured mammalian cells, and the process can be difficult and expensive. So making pharmaceutical proteins in transgenic farm animals — sometimes called 'pharming' — was hailed as a way to drive down production costs while boosting volume.

Sweetheart, for example, can be milked using standard dairy equipment; the protein is then purified and dried to a white powder. Maintaining her is a matter of feeding and milking. Need to scale up production? Breed more Sweethearts. The cost of using transgenic goats has been calculated to be anywhere from 3 to 100 times less than cell-culture processes.

Biopharmaceuticals is now a US\$33-billion

industry, but transgenic animals have yet to play any part in it and pharming companies have had a difficult history. In 2001, PPL Therapeutics of Edinburgh, UK — an offshoot of the group that created Dolly the cloned sheep — brought a pharming drug called alpha-1-antitrypsin to clinical trials. The drug, produced in sheep's milk, was meant to treat lung disorders. But reports of wheezing among patients prevented the drug from advancing to phase III trials. The company folded two years later. Its competitor, Pharming, nearly went broke around the same time, after deciding to switch from rabbits to cell culture for production of another drug. It has since stabilized as Pharming Group, based in Leiden, the Netherlands.

Animal crackers

The turmoil created a negative image of the transgenic animal market. "You've got one company running out of money and the other running into regulatory trouble," says Harry Meade, vice-president for research and development at GTC. "It certainly gave us pause. We were the last dog standing." GTC struggled too; its stock currently hovers around \$1.50 a share, down from a high of nearly \$50 in 2000.

Today, the pharming landscape is speckled with a few small companies. Venture capitalists have largely shied away from the technology, and bigger pharmaceutical companies have not embraced it either.

Pharming Group probably faces a tough road even if its latest product achieves regu-

"The industry has been holding its breath for over a decade."

latory approval, says Marcel Wijma, senior equity analyst at SNS Securities in Amsterdam. The company's drug, Rhucin, is targeted at people who have hereditary angioedema, a painful and sometimes fatal condition that causes severe swelling in various body tissues. The condition affects only 30,000 people in the Western world. Three other companies are also developing drugs for the condition; Pharming Group is the only one working with transgenic animals.

Even if Pharming wins 40% of the hereditary angioedema market, Wijma estimates that it would not make more than \$100 million to \$250 million a year. Couple this with the predicted \$50 million that GTC could make with ATryn, and drugs produced from transgenic animals would be a mere drop in the \$33-billion



ocean. Both companies are planning clinical trials for other uses of their drugs, however, which could expand their potential. Navdeep Jaikaria, a biotechnology analyst at Rodman & Renshaw in New York, suggests that GTC could make up to \$250 million a year from ATryn, if it were able to market it for a range of clotting disorders. The EMA approval currently covers its use for only very specific cases of a hereditary disorder.

And the expense of biopharmaceuticals means pharming drugs is still attractive, says Jaikaria. It costs \$10,000 a year, for instance, to treat an anaemic dialysis patient with Epogen, a recombinant form of human erythropoietin. Pressure from health-insurance companies, Jaikaria argues, will push the industry to take another look at farmyard pharmaceuticals.

Culture shock

Competing technologies have advanced in the meantime, however. Peter Knight, an analyst at Wood Mackenzie in Edinburgh, points to the pharmaceutical company Wyeth, based in New Jersey. Wyeth recently created a cell-culture system that can generate 9 grams of protein per litre of culture. Just four years ago, its standard was 1 gram per litre. So cell-culture yields are catching up with production rates from transgenic animals, which typically yield between 1 and 10 grams of protein per litre of milk.

"Pharming is certainly a technology that



GTC

The bleat goes on: a drug from transgenic goats has finally been approved, in Europe at least.

five years ago was attracting a huge amount of interest," says Knight. "But it does seem to have lost ground to other methods."

Competition is also provided by transgenic plants, Knight adds. Like animals, plants can be engineered to produce therapeutic antibodies and vaccines. Admittedly, it is harder to stop transgenic lines escaping into the environment, and there is a greater chance that proteins made in plants will provoke adverse immune responses in humans. But they promise shorter production times and a lower probability of contamination by viruses or prions.

Big companies, such as Indianapolis-based Dow AgroSciences and Swiss firm Syngenta, are already backing pharmaceutical production in plants. In fact, the next transgenic drug expected to gain EMEA approval is CaroRX, a product from California firm Planet Biotechnology that stops a cavity-causing bacterium from sticking to teeth.

Meanwhile, companies such as GTC and Pharming continue their slow march to the animal market. Interestingly, GTC's next big project is producing alpha-1-antitrypsin, the drug that caused PPL Therapeutics to fail.

Meade is undaunted by the drug's fraught history. "It's fair game," he says. "We've got goats that make 20 grams per litre." ■

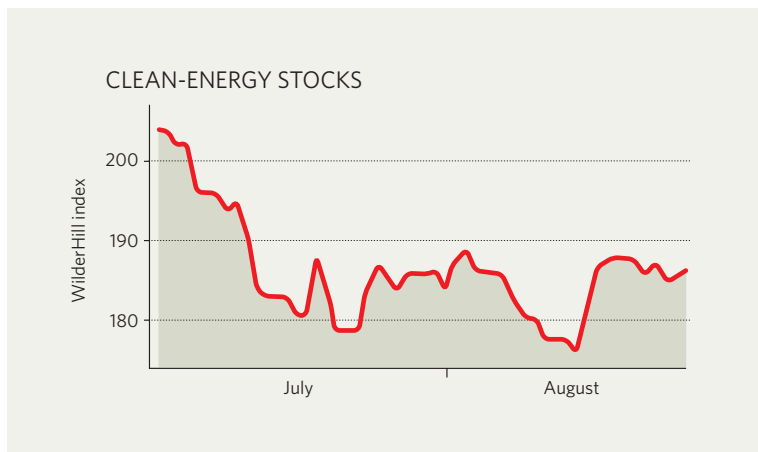
IN BRIEF

TOXIC TURNAROUND A US federal judge has overturned a 2004 rule that was meant to streamline the regulatory approval of pesticides. Manufacturers must now once again consult with officials from the Fish and Wildlife Service over the effects their chemicals might have on endangered animals and plants. A coalition of environmental groups that had sued the government hailed the decision on 24 August as a victory. In a separate ruling, a federal appeals court on 29 August upheld an earlier decision that the Environmental Protection Agency did not violate clean-air laws by allowing more of the pesticide methyl bromide to be produced.

BIG DEAL In one of the largest deals ever between a drug company and a biotechnology company, GlaxoSmithKline has signed a \$1.5 billion agreement with a northern California firm called ChemoCentryx. The firm will get \$63.5 million up front, with the rest of the money to follow if it meets certain targets while developing drugs. Its research will focus on four particular chemical receptors. The deal includes Traficet-EN, a drug to treat inflammatory bowel disease that is currently in clinical trials.

SPACE RACE Lockheed Martin of Bethesda, Maryland, has won a contract to build the next generation of spaceship for NASA. The deal is worth \$3.9 billion. The vehicle, dubbed Orion, will take astronauts to the international space station and make extended visits to the Moon. The company beat the combined forces of Northrop Grumman, headquartered in Los Angeles, and the Chicago-based Boeing to get the contract, which provides for the construction of two capsules: one for astronauts and a second to carry cargo to the space station. The contract sets aside \$3.5 billion more for additional capsules after the prototypes are built, as well as \$750 million for support and modifications as needed.

MARKET WATCH



Clean-energy stocks have continued to drop over the past two months, although analysts say the plunge is a natural stock-market correction rather than any sign that the industry is in trouble.

Companies that specialize in solar, wind, biofuels and other clean energy technologies continue to thrive overall — even though the WilderHill Clean Energy Index (ECO on the American Stock Exchange) has dropped markedly since May.

The fall reflects pessimism in the stock market as a whole, says Robert Wilder, chief executive of WilderShares in Encinitas, California, which maintains the index.

"The industry is catching up with a bulge of investment that happened over the past 18 months," adds Michael

Liebreich, chief executive officer of New Energy Finance in London.

Now it's up to companies to make good on all the recent investor interest — for instance, by increasing the manufacture of wind turbines or silicon for solar cells.

Public share offerings have quieted down through the summer, but several clean-energy companies — such as Hawkeye Holdings, an ethanol producer based in Iowa — are expected to go public soon.

Liebreich says that clean-energy stocks should begin to rise again before the end of the year, as growth takes off again after the correction. "Directionally there's no question where it's going," he says.

Alexandra Witze

SOURCE: WILDERSHARES



A NEW DAY DAWNING?

Sunlight is a ubiquitous form of energy, but not as yet an economic one. In the first of two features, **Oliver Morton** looks at how interest in photovoltaic research is heating up in California's Silicon Valley. In the second, **Carina Dennis** talks to Australian researchers hoping to harness the dawn Sun's heat.

Silicon Valley sunrise

The Sun provides Earth with as much energy every hour as human civilization uses every year. If you are a solar-energy enthusiast, that says it all. No other energy supply could conceivably be as plentiful as the 120,000 terawatts the Sun provides ceaselessly and unbidden. If the tiniest fraction of that sunlight were to be captured by photovoltaic cells that turn it straight into electricity, there would be no need to emit any greenhouse gases from any power plant.

Thanks to green thoughts like that, and to generous subsidies from governments in Japan and Germany, the solar-cell market has been growing on average by a heady 31% a year for the past decade (see chart, overleaf). One of the most bullish industry analysts, Michael Rogol, sees the industry increasing from about US\$12 billion in 2005 to as much as \$70 billion in 2010. Although not everyone predicts such impressive growth, a 20–25% annual rise is widely expected. The market for shares in solar-energy companies is correspondingly buoyant.

And yet in the projections of energy supply made by policy analysts and climate wonks, solar remains so marginal as to be barely on the map at all. At the moment, the world's total

installed solar cells have a capacity of about five gigawatts. That looks small compared with almost 400 gigawatts for nuclear power and much more than 1,000 gigawatts for coal. And that's before taking into account the fact that solar cells do not produce electricity at their peak rating all the time. Even within the world of renewables, solar is dwarfed by wind power and hydroelectricity, simply because the technology is much more expensive. And expert opinion does not expect growth in the field to change the picture very much: a 25% annual growth in installed capacity for the next 15 years would still see solar photovoltaics producing just 1% of the world's energy.

Points of view

Reconciling the solar-cell industry's optimism with global indifference is basically a matter of perspective. Seen from the viewpoint of a small industry, solar's recent decade of expansion is indeed extraordinary. But even heady growth is not enough to spur a radical overhaul of energy infrastructure when you start such a long way behind your competitors. So, although no one doubts that solar electricity will become cheaper in the future,

D. SANGER PHOTOGRAPHY/LAMY



Lighting the way: Chris Eberspacher has developed a relatively low-cost way to manufacture solar cells.

few expect it to do so fast enough to force radical change.

Few worldwide, that is. In California's Silicon Valley, the corridor of land along the southwest side of San Francisco Bay, the outlook is more optimistic. Home first to the semiconductor boom and then to the Internet boom, the valley is perhaps the most fertile environment for new technologies in the world. As well as an extraordinary density of successful technology-based companies, it boasts world-class research universities, abundant capital, and a cultural fixation on getting to the future first and making money from it. The dot.com bust of 2000 did relatively little to dent the valley's fundamental strengths and attitudes; instead, it left the area's entrepreneurs and venture capitalists looking for somewhere else to put their millions.

'Cleantech' of all sorts, from water purification to biofuels, is currently the place they want to be. In 2005, \$1.6 billion in venture capital went into cleantech, a growth of 35% year on year according to the Cleantech Venture Network, an umbrella group. The high-profile venture-capital firm Kleiner Perkins Caulfield and Byers is putting \$100 million of its latest \$600-million investment fund into cleantech start-ups. Bill Joy, a partner at Kleiner Perkins who used to be chief scientist at Sun Microsystems, says that the firm will look at perhaps 1,000 different cleantech ideas in the next year. And amid all these opportunities, photovoltaics seem to resonate most with Silicon Valley's history and culture.

One attraction is technological familiarity. Solar power has grown up in the shadow of the chip industry, using its cast-off materials and technologies. The silicon in traditional solar cells comes from the same suppliers who feed the chip market; new techniques to make solar cells often use processing technology, such as chemical vapour deposition, that is already widely used in the production of integrated circuits. Miasolé, a Silicon Valley solar start-up in which Kleiner Perkins has invested, uses expertise derived from the manufacture of computer hard drives.

But there is a broader cultural attraction, too. The potential of solar power to decentralize energy generation — a potential shared, to a lesser extent, by wind power — appeals to a culture that places huge societal significance on the empowering spread of the Internet. And a business community that saw personal

computers go from hobbyists' workshops to almost a billion of the world's desks in 30 years is not fazed by the small size of the solar market today, but energized by the possibilities of tomorrow.

It's also a help that Silicon Valley is sunny not just in its outlook; a solar cell in California can produce almost twice as much electricity a year as one in the Ruhr.

Catching the rays

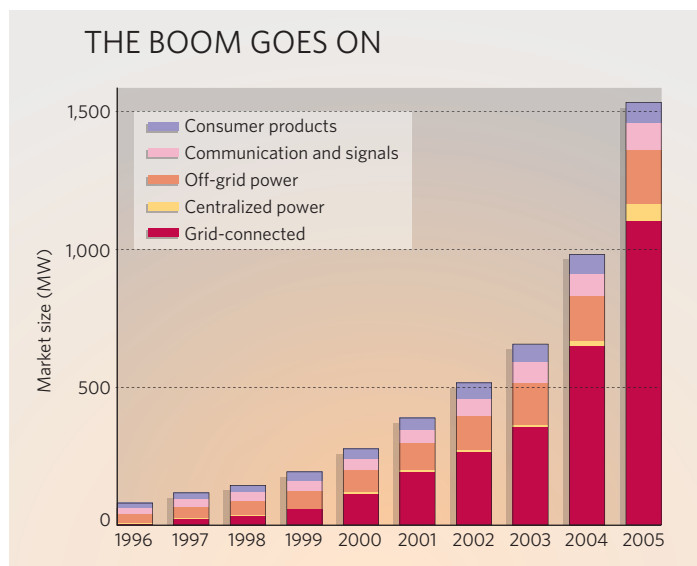
The poster child for Silicon Valley's interest in solar power is a company called Nanosolar, based in a distinctly unimposing one-storey building next to Palo Alto's municipal airfield. Disappointingly, it has no solar panels on its roof, although there is a smattering of Prius hybrids in its parking lot.

Nanosolar was founded in 2001 by Brian Sager, a biotech veteran with expertise in intellectual property, and Martin Roscheisen, an Austrian entrepreneur who mixes grandiloquence, enthusiasm and edginess. Like many Silicon Valley entrepreneurs, Roscheisen had a good track record: companies he had had a founding role in had sold for more than a billion dollars. And Nanosolar quickly attracted 'angel' investors with powerful reputations, including the founders of Google. In 2002, it became the first solar company to raise money on Sand Hill Road, Palo Alto's superconcentration of venture-capital firms.

Nanosolar was not founded with one specific solar technology in mind, says Roscheisen. That's just as well, because the range of technologies it spent its first years investigating have not as yet panned out. The 'nano' in the company's name reflected an early belief that the use of very small structures would allow novel photovoltaics made of organic molecules to overcome certain difficulties, such as being able to transfer charge only over very short distances. But the founders soon concluded that "the organic part was going to require a number of years to mature", says Chris Eberspacher, the company's vice-president of engineering. "And even when mature it would not be very efficient, and not be very durable."

That was where Eberspacher came in. He had been a solar aficionado since a school trip to the University of Texas, Austin; instead of being awestruck by the Texas Turbulent Tokamak fusion experiment being shown to potential students, he found himself drawn to the ramshackle alternative-technologies centre across the street. Eberspacher went on to become head of research and development at Arco, once the largest maker of solar cells in the world, before starting a company of his own.

When Nanosolar had trouble getting its organic technologies



W. BREUER to work, the company leaders looked around for something less risky that they could get to work in the medium term. They had capital; Eberspacher had technologies that, while still innovative, were considerably more tried and tested than those Nanosolar had been working on. He licensed the technologies to them and joined up in 2005.

That technology is now being scaled up for production at Nanosolar's first factory, which aims to produce more than 200 megawatts of solar cells in its first year and 430 megawatts a year later. That would make it among the largest solar-cell fabrication facilities in the world, and by far the largest devoted to this sort of 'thin-film' solar cell.

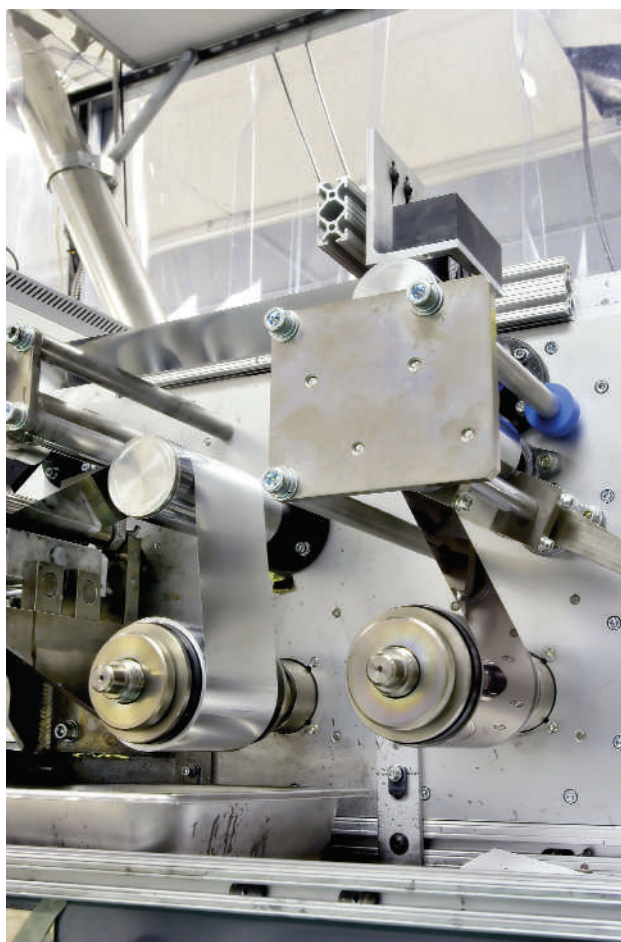
Traditional silicon solar cells are made out of chunks of silicon 200 micrometres thick or more, but slivers a single micrometre across can suffice for a 'CIGS' thin film. CIGS cells are made up of copper, indium, gallium and selenium. Even though some of those elements are increasingly expensive — the price of copper has more than tripled in the past four years and the price of indium has shot up by a factor of ten — they are used in such sparing amounts that this is not too great a problem.

What is a problem is that making very thin layers of CIGS has often been a complicated and expensive business, typically involving carefully controlled vapours being laid on to surfaces kept in vacuums. "The silicon [photovoltaic] industry got founded on a wafer technology we inherited from the integrated-circuit industry," says Eberspacher, "and the thin-film photovoltaic industry got founded on deposition techniques inherited from them in the same way." But expenses that are reasonable for materials that process information are too high for materials that process energy.

On a roll

Decades of development have made CIGS cells as efficient as mass-market silicon cells; they can convert about 15% of incoming solar radiation into outgoing electrical current. They are also durable — the National Renewable Energy Laboratory in Golden, Colorado, has been running some since 1988 without any significant degradation. But they are not yet cheap to produce.

The leaders of Nanosolar think that Eberspacher's techniques offer a way around that. They make tiny CIGS particles and mix them into a sort of ink, printing them on to a substrate of metal foil and curing that foil in such a way that the particles condense into a continuous semiconductor. The cells should hit the market in 2007, and the fact that the particles



Thin-film solar cells can be made continuously using a roll printer.

are nanometres across means that the company is still accurately named — although more by luck than judgement.

Perhaps the most attractive aspect of the Nanosolar process is that it can be carried out on foil being continuously pulled off one roll on to another, allowing very high throughput. Such 'roll-to-roll' technologies make it possible to build a large factory with a relatively small investment and make cells cheaply. Roscheisen boasts that the production costs for Nanosolar's cells are so low that even if you subtracted the costs of materials, manpower and energy from a traditional silicon factory, its cells would still be more expensive than those of Nanosolar.

Whether Nanosolar can live up to that boast remains to be seen. The fact that it has just raised a further \$75 million in private capital suggests that some fairly serious investors believe it will. Whether or not it succeeds, many other companies are trying the same thing. Miasolé in Santa Clara is starting to produce CIGS films made with its hard-drive technology. Earlier this year, Applied Materials — a far larger Santa Clara company that sells the machinery used to make chips and flat-

panel displays — acquired the rights to ways of making thin-film silicon-based photovoltaics developed in Germany. And Nanosys, a Palo Alto firm, is working on nanostructures that could minimize current difficulties with the sort of organic polymer-based solar cells that Nanosolar was looking at in its early days.

The valley does not have a monopoly on innovation. DayStar, based in Halfmoon, New York, is also pursuing CIGS thin films, as is Wurth Solar in Germany. In Austin, Texas, B. J. Stanbery has founded a company called Heliovolta. Stanbery, who has been working with CIGS thin films since the early 1980s, when they were first under development at Boeing, has developed techniques for printing such films on a variety of substrates, speeding up their manufacture. In Lowell, Massachusetts, a company called Konarka is working on a novel system for using dyes to produce solar power from flexible plastics. The interaction of sunlight with these organic dyes produces solar power in a way that is perhaps more similar to photosynthesis than to the semiconductor processes in normal solar cells.

Measuring up

But even if one or more of these companies manages to make solar cells a great deal cheaper, it will be only the beginning. Manufacturing the cells accounts for just half of the roughly \$6 per watt it costs to get a solar-cell system up and running. The remaining cost is needed to put them into a protective, mountable module, tune their output from direct current to alternating current, and install them.

This has various implications. One is that cells below about 10% efficiency have a hard time making economic sense, because the costs of mounting and installing cells in traditional models get bigger the larger the area involved, and low-efficiency cells require larger areas. Another is that even if you gave away 15%-efficient cells for free, systems using modules such as today's would still be too expensive for many applications. This is why Nanosolar and almost all

"The business community is not fazed by the small size of the solar market today — it's energized by the possibilities of tomorrow."

the other recent solar start-ups take a strong interest in new ways of mounting their cells — ways that take advantage of their light weight or flexibility. Eberspacher hopes, for example, that such light-weight systems could be used on Nanosolar's own roof, which is too flimsy to take the load from a traditional array.

The ultimate aim, says Stanbery, is to integrate the cells straight into building materials

of all sorts. New houses, he points out, need roofs anyway. Photovoltaic tiles could be wired into the house from the start. "Integrating the photovoltaics as a coating," he says, "is frankly the only practical and cost-effective way to do it." Heliovolt's printing process is meant to help make that integration possible. And Konarka talks of adding its dye-based 'Power Plastic' to more or less anything, from windows (where it would just cream off a bit of the light) to wind sheeters.

None of these technologies, however cleverly mounted, will get the costs of generating electricity low enough for solar power to compete directly with coal, gas, wind or nuclear. But because solar panels are inherently easily decentralized, they do not have to compete with the cost of generating electricity; they just have to compete with the price consumers pay for it. This is four or five times more than the cost of generation, because the power companies need to pay for transmission networks, build new plants and please shareholders.

So the industry's aim is to get significantly below 'grid parity'. This is the point at which the cost of borrowing the money to buy and install a solar-power system is more than covered by savings on your electricity bill. At the moment, grid parity is not quite within reach; in most places with a lot of solar cells there is or has been a great deal of government subsidy. In Germany, a particularly powerful subsidy is a government requirement that electric utilities be willing to buy electricity generated by small photovoltaic installations, such as those in homes and small businesses, at more than 50 cents a kilowatt-hour. Although this rate decreases over

time, it is still a costly subsidy, and some wonder how long it can last in its present form. In its favour is popularity with the electorate — and, of course, with Germany's producers of solar cells.

Reaching grid parity is not in itself enough. But if a mixture of much cheaper cells and adaptable, easily installed modules could bring the total cost of installation down by a factor of three, solar energy would start to look prudent, analysts say.

Spending to save

There is a lively research agenda in basic materials science that a thriving solar industry could use to drive costs down further still. Michael McGehee of Stanford University, the young investigator whose work on organic photovoltaics was part of the original inspiration for Nanosolar, is developing a proposal for a major initiative in solar energy research. "We are going to apply for a large centre here, funded by the Department of Energy," he says. "We have a team of people who will work on using sunlight to excite a semiconductor, or to split water to make hydrogen." If fully funded, it would have 16–20 principal investigators and be one of the biggest research groups with a specific target at Stanford.

Steve Chu, the Nobel prizewinning Stanford physicist who now runs the Lawrence Berkeley National Laboratory on the other side of San Francisco Bay, has ambitious plans for an initiative called Helios. This would integrate new photovoltaic research with studies into other ways of capturing and storing sunlight, such

as biofuels; the idea is currently under review at the Department of Energy.

Also testimony to the research interest in this area is the way that it is being presented at meetings. Organizing a session on solar applications at last November's meeting of the Materials Research Society, McGehee found himself swamped with hundreds of abstracts; it

was the third most popular of the meeting's 40 sessions.

One potential source of funding for all this research is Proposition 87, which will be on the California ballot this November. The proposition, which is strongly opposed by oil producers, would increase the cost of drilling fees in order

to raise \$4 billion for clean-energy initiatives, including research. Vinod Khosla, a former partner at Kleiner Perkins, is a main backer of the initiative, and other venture capitalists are also on board.

But the fact that Silicon Valley is abuzz with solar enthusiasm doesn't necessarily mean that all the activity there will trigger a revolution. Someone elsewhere might come up with the key breakthrough technology. And just because venture capitalists are successful in making money doesn't mean they will effect major economic changes. As Stephen Levy of the Center for the Continuing Study of the California Economy points out, venture capitalists have been saying that biotech would be the big new growth sector for years, "and it is still 'just about to explode', with an emphasis on the 'just about'."

Solar enthusiasts can respond that solar cells have no Food and Drug Administration to face, and that they don't need to invest hundreds of millions to get a product to market, as drug developers do. Yet a decade's growth, however buoyant, doesn't by itself mean that much. That growth needs to last for several decades to change an economy, and needs to accelerate to an even higher level to change the world.

The difference between growing at a more than respectable 25% a year and at 44% a year — the rate at which volume grew in 2005 — is the difference between doubling in size in just over three years and in just over two. That may not sound a great deal, but over 15 years it means something growing at 44% would outdo something growing at 25% by a factor of eight. Between now and 2050, the difference is a factor of 500. And that could be the difference between providing just 2% of Earth's energy needs — and 10 times those needs.

The remarkable thing is that the products of the semiconductor industry have grown at a yet faster rate for a similar length of time. If Silicon Valley can apply Moore's law to the capture of sunshine, it could change the world again. ■

Oliver Morton is *Nature's* chief news and features editor.

See Editorial, page 1.

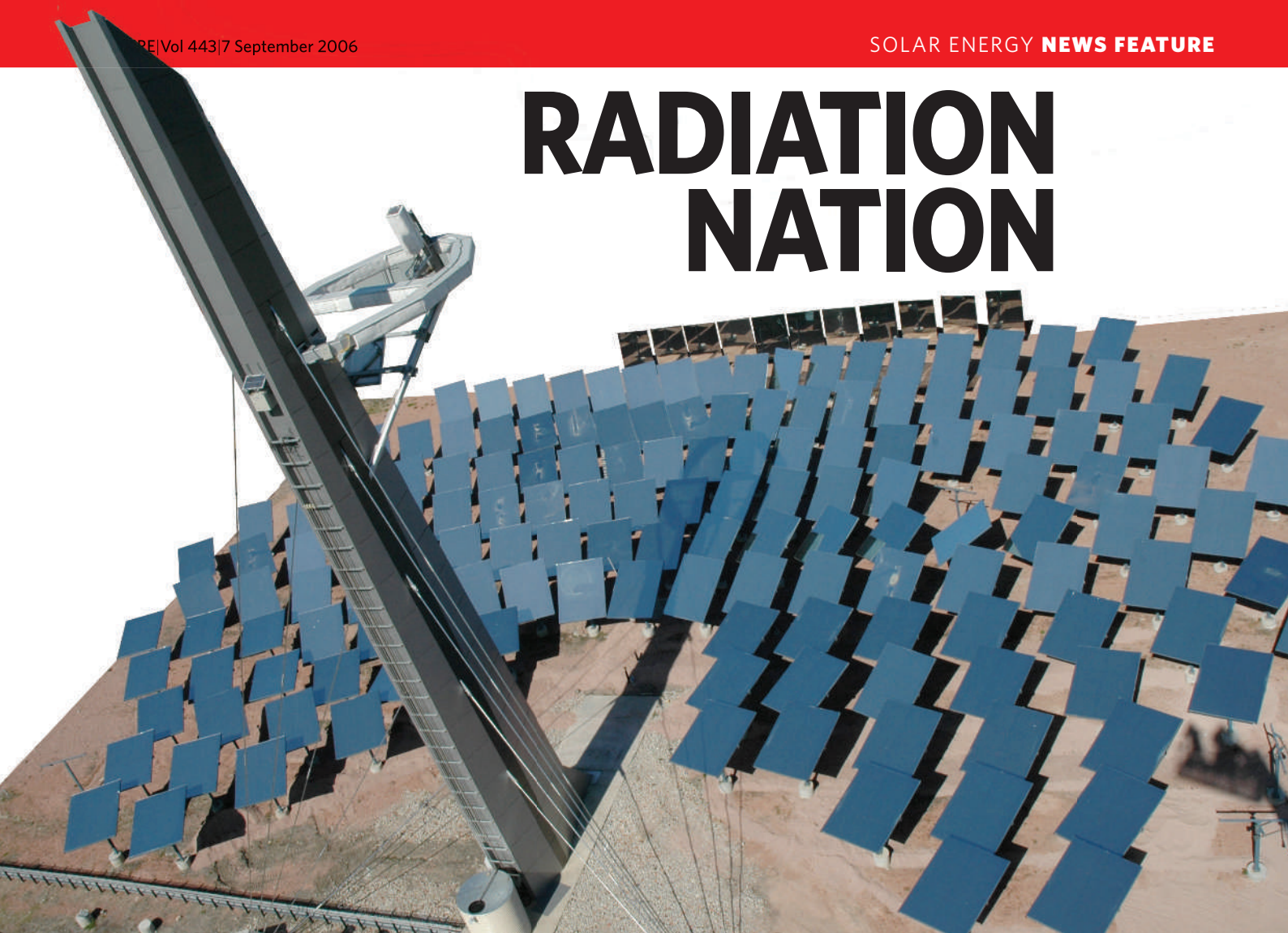
"Integrating the photovoltaics is frankly the only practical and cost-effective way to do it."

— B. J. Stanbery



High flyers: the bright sparks of Silicon Valley can see a future in solar cells.

RADIATION NATION



Australia receives more solar radiation per square metre, on average, than any other continent. Although turning this radiation into electricity is one option, another is finding ways to make use of its heat. We spoke to Australian proponents of two very different solar-thermal systems, both rather confusingly known as solar towers: Roger Davey, executive chairman and chief executive of EnviroMission, based in Melbourne; and Wes Stein, from the Energy Technology Division of the Commonwealth Scientific and Industrial Research Organisation, based in Newcastle.

How does your solar-thermal project work?

Roger Davey: In our solar tower, solar radiation is used to heat the air captured under a large greenhouse. The roof of the greenhouse is sloped towards the centre, where there are turbines and a very tall tower that creates a chimney effect. Hot air rises up and out of the tower, rushing through turbines at the base of the tower to produce power.

Wes Stein: We have 200 small mirrors that track the Sun during the day and concentrate its rays onto a single point on the tower. This provides temperatures hot enough to drive a chem-

ical 'reforming' reaction. Methane and steam go in, energy and a catalyst are added; the resulting gas [carbon monoxide and hydrogen] is called syngas in the industry — we call it SolarGas. This gas has 26% more energy per kilogram than methane — it is solar energy embedded in chemical bonds. So the plant improves the energy value of the gas, and provides a precursor for other energy products, such as liquid fuels. We could link this process with carbon burial so that the carbon dioxide from the fossil-fuel component was sequestered.

How big is your solar tower?

RD: The original concept design made it 1 kilometre high and 7 kilometres in diameter, but we have developed new technologies that will make it substantially smaller and more powerful. A 50-megawatt demonstration plant will be built at Tapio Station, about 22 kilometres northeast of Mildura in New South Wales. This is not necessarily the optimum size, but it is the optimum size to build in the first instance to show the robustness of the technology. The front-end engineering and design are currently under way, and it would be premature of me to talk heights and sizes now, and then alter them in a week's time.

WS: The tower reactor is 15 metres above the mirrors, and the tops of the mirrors are about

These mirrors in the Australian landscape focus the Sun's rays to power the production of 'SolarGas'.

3 metres off the ground. The total ground area is 40×40 metres, but that is because of our site constraints — it doesn't necessarily represent the ideal module size. So I don't want to make a big deal of the specific dimensions. It will produce up to 250 kilowatts of electricity. However, this is a research-size system. We designed this tower to represent a single module that could be replicated again and again to make it up to any desired capacity.

How does your system fit into the existing energy infrastructure?

RD: It operates as a power plant and feeds into the normal electricity grid.

WS: The gas can be used in gas turbines to make electricity or we can turn it into liquid transport fuels. We are investigating the use of existing gas pipelines to transport SolarGas. One idea is to do the solar-reforming reaction out in the desert region, where many gas pipelines originate, putting the SolarGas into the pipeline and transporting it downstream. We can have all the convenience of gas but with all the benefits of solar energy.

SolarGas can make liquid transport fuels and also fertilizer. After power plants, transport

W. STEIN/CSIRO

W. STEIN/CSIRO



"We can use SolarGas in a broad range of sectors, not just electricity."
— Wes Stein

and agriculture are the biggest greenhouse-gas emitters in Australia, so it gives us an opportunity to target more than electricity.

How are you trying to improve the technology?

RD: We can now create far greater heat under the collector roof. It captures more heat and retains more heat — it's a bit like double-glazing.

We now also have a method of storing heat in brine ponds, so that on days of lower solar radiation we can bring in that stored heat to create the temperatures we need. That means we have a system that could operate 24 hours a day, 7 days a week, 365 days a year.

WS: We believe cost reductions will come from improving the efficiency of the reaction and reducing the cost of the mirror field through mass production. We have two areas of research we are working on. One is to reduce the reaction temperature to around 550 °C [from around 800 °C]. That greatly reduces the capital costs of the mirror field, because fewer mirrors are required. We are using a novel arrangement of membranes to allow a much lower reaction temperature. We are in the middle of a patent application, so that's all I can say at present.

The second concept uses carbon dioxide as the reforming agent rather than steam, so we would be using a waste stream. Coal-seam methane is a rapidly growing resource in Australia. Methane from coal seams comes out with a fair amount of CO₂. Normally, that CO₂ is stripped off before the gas goes into the downstream pipeline — so we would make use of that discarded CO₂. But we need to develop new catalysts for that reaction.

How much does it cost?

RD: The original-concept 200-megawatt plant would have cost around Aus\$800 million (US\$610 million). The 50-megawatt plant will be substantially less, owing to a reduction in the structural dimensions, with final costings anticipated at the completion of the front-end engineering and design. We are working on the engineering of the optimally sized power

plant with the optimum output, incorporating the new technologies. I could guess, but it's better to be sure than to mislead.

WS: This solar tower has cost us Aus\$1.5 million to build. But we think that if we apply the same learning curve that turbines have experienced over the past 15–20 years, this same system would cost us around Aus\$300,000 to build in the future.

What are your advantages over rival solar technologies?

RD: We can guarantee output to meet the demand, just like coal-fired plants.

WS: The advantages are that we can use SolarGas in a broader range of sectors, not just electricity; the very-low-cost potential; and the fact that it overcomes the transport and storage issues of solar energy.

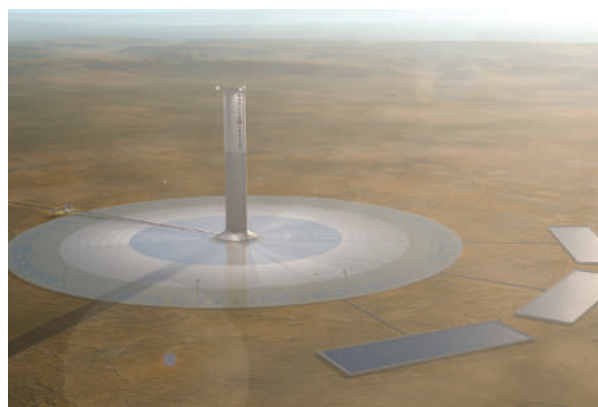
Solar thermal has been around for a while — why has it not caught on?

RD: I think it has caught on. That's why we're here.

WS: One of the difficulties it has had is the concept that it has to be built large for cost reductions to occur. It's one thing to build a system at the small demonstration level and enthuse everyone with that, and then say 'now I need to build 100,000 of these to be cost-effective in a 50-megawatt system'. I think the jump from that small to large scale has been too much for investors to handle, even though the modelling and calculations have always shown that it would be cost-effective.

Worldwide, investment in solar-thermal research is fairly low; why do you think that is?

RD: If you take the solar tower as an example, the original design had to be big to produce



Artist's impression of EnviroMission's solar tower, which will harness the Sun's heat to generate electricity.

enough power to make enough money to pay for the capital costs. But with the additional technology we have added, we have created far more power out of a smaller plant.

WS: It is because of the difficulty investors have in biting off big chunks compared with the easier task of biting off small chunks. Photovoltaics are very sexy — they don't move, they

"We have a system that could operate 24 hours a day, 7 days a week, 365 days a year."
— Roger Davey



J. SMITH/AAP IMAGE

don't make any noise sitting on top of a roof, and they can be made to look good on buildings. They are still very expensive, but they come in small chunks. It is much easier for investors to go with that concept than with a one-off, large, solar-thermal power station. It's only now, as a result of initiatives and incentives around the world, that solar-thermal technology is starting to take off globally.

What stops power companies from making big investments in solar thermal?

RD: Power companies have investments in other assets, don't they?

WS: Purely cost. At the end of the day, it comes down to cost of the technology.

How much of Australia's electricity needs could solar thermal provide in 20 years — or in 50?

RD: We plan to develop approximately 2,100 megawatts of installed capacity by 2020. By 2030, we plan to have installed in excess of 6,500 megawatts of power, which would be the equivalent of the electricity usage of about 10 million households based on current energy use.

WS: I see no reason why, by 2050, solar energy couldn't be supplying at least 25% of Australia's energy mix.

Is your technology really only suitable for countries with big, empty deserts?

RD: No. Our solar tower operates on temperature differentiation — the creation of an environment with strong differentiation to ambient temperature will regulate where development is able to occur. Unfortunately, big, empty deserts do not serve electricity grids and have inherent supply-chain issues.

WS: Sensibly, yes. Technically, there is no reason why you can't do it in lower solar regions but the cost will be higher. I'm not advocating solar being a single solution for all Australia's energy needs. Theoretically it could be, but practically I don't think that will happen. ■

Carina Dennis is Nature's Australasian correspondent.

Refund fertility-treatment costs for donated embryos

SIR — Your recent Editorial, Special Report and Commentary article (*Nature* **442**, 601, 606–608 and 629–630; 2006) focused on the ethics of payments to donors of human eggs for research purposes. There is, however, a source of embryos that would not have been generated specifically for this purpose.

The increasing popularity of clinically assisted reproduction has led to an accumulated surplus of frozen embryos in fertility clinics worldwide. Couples who have attained success in clinically assisted reproduction are often faced with the dilemma of what to do with their surplus frozen embryos: whether to discard them, donate them for research or give them to other infertile couples. When such embryos are donated for research, a pertinent question is whether the donor should receive payment, at least to the extent of reimbursement of storage and medical fees.

Clinically assisted reproduction is an expensive medical procedure that imposes a heavy economic burden on sub-fertile couples in countries where the patient pays. Patients undergoing fertility treatment often use their entire savings for the purpose, leaving scant resources for childcare and education. Hence, it seems fair and ethically justifiable for them to receive reimbursement for part of the medical and storage fees that they had previously paid for their treatment, if they later decided to donate their surplus frozen embryos for stem-cell research.

The portion of medical fees most eligible for reimbursement would be the amount of money directly spent on cryopreservation and storage of surplus embryos in liquid nitrogen. Another eligible component is the *pro rata* cost of superovulatory drugs, which makes up a substantial portion of the medical fees, particularly in the case of highly purified recombinant gonadotropins. For example, suppose that ten embryos are produced during a prospective donor couple's fertility treatment and six of these are used for treatment. If the remaining four embryos are cryopreserved, stored under liquid nitrogen and subsequently donated for stem-cell research, then the fraction of drug prescription costs reimbursed to the patient would be 40%. The reimbursement could come either from the research grant *per se*, or from another charity or trust supporting stem-cell research. In this manner, former patients who had previously paid hefty medical bills for fertility treatment would receive an influx of much-needed cash for raising their new family.

The chief ethical objection would be the perception of reimbursement being equivalent to the sale and purchase of the human embryo as a marketable commodity.

Nevertheless, it can easily be counter-argued that the payment does not involve embryos donated specifically for research, as patients would be reimbursed only for money that they had already spent in generating surplus frozen embryos in case of initial failure in attempting clinically assisted reproduction. Because patients have to qualify for assisted-reproduction programmes, the issue of 'undue incentive' for egg donation does not arise. Of course, it is essential that patients should receive professional counselling so that they can make a well-informed decision, before any reimbursement is given.

Boon Chin Heng, Tong Cao

Stem Cell Laboratory, National University of Singapore, 5 Lower Kent Ridge Road, 119074 Singapore

Dogma, not faith, is the barrier to scientific enquiry

SIR — Your News story "Genomics luminary weighs in on US faith debate" (*Nature* **442**, 114–115; 2006) reports that the Christian molecular biologist Francis Collins embraces 'theistic evolution' to reconcile scientific facts and religious views.

In his book *The Language of God* (Free Press, 2006), Collins discusses the ideas of Theodosius Dobzhansky, a darwinist and the main architect of the modern synthetic theory of biological evolution. In a famous article, "Nothing in biology makes sense except in the light of evolution" (*Am. Biol. Teach.* **35**, 125–129; 1973), Dobzhansky described his religious beliefs: "It is wrong to hold creation and evolution as mutually exclusive alternatives. I am a creationist and an evolutionist. Evolution is God's, or Nature's, method of Creation."

In contrast to modern creationists, Dobzhansky accepted macroevolution and the documented age of Earth. He argued that "the Creator has created the living world not by caprice (supernatural fiat) but by evolution propelled by natural selection".

He collaborated for many years with Ernst Mayr, who, when asked about his religious views, replied: "I am an atheist. There is nothing that supports the idea of a personal God. On the other hand, famous evolutionists such as Dobzhansky were firm believers in a personal God. He would work as a scientist all week and then on Sunday get down on his knees and pray to God" (*Skeptic* **8**, 76–82; 2000).

In about 1950, Dobzhansky and Mayr founded our modern 'atheistic' evolutionary theory. Their work showed that Christians and atheists can cooperate to develop scientific theories, as long as religious dogma is not mixed up with facts and experimental data. Unfortunately, this is exactly what young-Earth creationists and intelligent-

design theorists are doing. They should read the 1973 essay in which Dobzhansky — an open-minded, non-dogmatic theist — thoroughly refuted their irrational claims.

U. Kutschera

Institute of Biology, University of Kassel, Heinrich-Plett-Strasse 40, D-34109 Kassel, Germany

Author lists: specify who did what to aid assessment

SIR — Deciding whether particular contributors deserve to have their names published in a scientific paper and in what order contributors' names should appear is not a simple issue, as highlighted in Correspondence by William F. Laurance ("Second thoughts on who goes where in author lists" *Nature* **442**, 26; 2006).

Nature's formatting guide advises authors to include a "statement to specify the contributions of each co-author" (www.nature.com/nature/authors/gta). If authors of all papers in all journals followed *Nature's* advice, their contribution would be clear to any reader, irrespective of order of names in the authorship lists.

This information could then be used as a guideline for assessing ranks of scientific productivity by mechanisms such as the UK Research Assessment Exercise. Such organizations would need rules for giving credit for each kind of contribution, which would not be an easy task. Nevertheless, in that the authors themselves have agreed their individual contributions to a published paper, this process should be easier than trying to institute a global rule for authorship order that can fairly be applied to many different disciplines.

I am the only contributor for the development and organization of the ideas within this Correspondence.

Marcos V. C. Vital

Universidade Federal de Viçosa, Departamento de Biologia Geral, 36570-000 Viçosa, Minas Gerais, Brazil

And finally...

SIR — We read with interest William F. Laurance's Correspondence letter "Second thoughts on who goes where in author lists" (*Nature* **442**, 26; 2006). We have also noted in this and other journals an increasing number of papers with joint first authors. Perhaps the time has come to accept joint last authors?

R. Martyn Bracewell, Steve Tipper*, Robert Rafal*

Wolfson Centre for Clinical and Cognitive Neuroscience, University of Wales, Bangor LL57 2AS, UK

*These authors contributed equally minimal amounts to this letter

COMMENTARY

Selling out on nature

With scant evidence that market-based conservation works, argues **Douglas J. McCauley**, the time is ripe for returning to the protection of nature for nature's sake.

Probably the most important trend in conservation science at the moment is 'ecosystem services', typically seen as economic benefits provided by natural ecosystems¹. They form the basis of most market-oriented mechanisms for conservation. The underlying assumption is that if scientists can identify ecosystem services, quantify their economic value, and ultimately bring conservation more in synchrony with market ideologies², then the decision-makers will recognize the folly of environmental destruction and work to safeguard nature.

But market-based mechanisms for conservation are not a panacea for our current conservation ills. If we mean to make significant and long-lasting gains in conservation, we must strongly assert the primacy of ethics and aesthetics in conservation. We must act quickly to redirect much of the effort now being devoted to the commodification of nature back towards instilling a love for nature in more people.

Gold rush

The proponents of market-based mechanisms for conservation bolster their argument by repeatedly citing one example: the Catskill/Delaware Watershed. Through this project, New York City invested in conserving a watershed that filters its water as effectively as a filtration plant, and more cheaply.

A growing number of ecologists, economists and environmental scientists hold this shining example aloft and proclaim that where there



L.B. AUPPY/GETTY IMAGES

The incentive to conserve the Catskill watershed could be lost if technology replaces natural filtration.

is one golden nugget, there must be others. They describe, mostly in hypothetical terms, a world of win-win scenarios. It is a message with broad appeal: for the public, which is notoriously averse to bad news; for business-oriented politicians, who see an opportunity to further liberalize markets while appeasing the environmentally anxious; for philanthropists who wish to do good without straying too far from their economic comfort zones; and for foundations that want to use the familiar capitalist rhetoric of ecosystem services to draw out new or wary donors.

It is both true and obvious that 'ecosystems', in some sense of the word, are necessary for human survival. It is also true that there will be cases in which it will be lucrative to protect nature, and that people will derive benefits from this conservation effort. However, ecosystem services are rapidly assuming an importance in discussions on conservation that is far out of proportion to their actual utility.

As conservation tools, ecosystem services are limited in four fundamental ways. First, the logic of ecosystem-service-based conservation rests on the implicit assumption that the biosphere is benevolent — that it provides us with useful services and protects us from malevolent

abiotic forces such as hurricanes, floods and rising temperatures. This reasoning ignores basic ecology: environments don't act for the benefit of any single species. There are myriad examples of what might be labelled 'ecosystem disservices'. Trees take water out of watersheds³; forests may be contributing to global temperature increases⁴; wild animals kill people and destroy property⁵; and wetlands can increase the risk of disease⁶. Market-based conservation strategies, as currently articulated, offer little guidance on how we are to protect the chunks of nature that conflict with our interests or preserve the perhaps far more numerous pieces of nature that neither help nor harm us.

Markets in flux

Second, although most conservationists would argue that nature should be conserved in perpetuity, the strength and direction of market forces that are now being called upon to motivate nature conservation are anything but perpetual. The often illusory and ephemeral relationship of the market to conservation is well illustrated by the case of a former coffee plantation, Finca Santa Fe, in the Valle del General of Costa Rica⁷. A recent study found that native bees from two forest fragments



A. GANDOLFI/NATURE PL

Profit-oriented conservation strategies may fail to protect animals that conflict with our interests.

adjacent to Finca Santa Fe yielded approximately US\$60,000 a year in pollination services to the coffee plants. This was hailed as an example of how conservation can yield 'double benefits' for biodiversity and agriculture.

Shortly after the conclusion of the study, however, Finca Santa Fe, probably affected by one of the worst dips in coffee prices this century, cleared its coffee and planted pineapple instead. Pollinators are irrelevant to pineapple production. So simple logic suggests that over a period of several years, the monetary value of the pollinators in forest fragments around Finca Santa Fe dropped from \$60,000 per year to zero.

To make ecosystem services the foundation of our conservation strategies is to imply — intentionally or otherwise — that nature is only worth conserving when it is, or can be made, profitable. The risk in advocating this position is that we might be taken at our word. Then, if there is a 'devaluation' of nature, as in the case of Finca Santa Fe, what are we to tell local stewards who have invested in our ideology, and how can we protect nature from liquidation?

Watershed down

Third, conservation based on ecosystem services commits the folly of betting against human ingenuity. The entire history of technology and human 'progress' is one of producing artificial substitutes for what we once obtained from nature, or domesticating once-natural services. One of the primary selling points for protecting the Catskill/Delaware Watershed was that the costs associated with constructing and operating a filtration plant would have driven up water prices in New York City. However, recent reports⁸ indicate that increased turbidity might ultimately force New York to turn to technology to filter its water, in essence negating this much-ballyhooed economic incentive for conservation.

Several other major US cities still rely on natural filtration, and in some of these cases it is difficult to imagine that technology will soon produce a cheaper artificial alternative to these natural watersheds. Yet it would also once have been difficult to imagine cost-effective manufactured alternatives to rubber and timber. Although we will never replicate all of the 'services' offered by nature, I would argue that conservation plans that underestimate the technological prowess of humans are bound to have short lifespans.

Lastly, although it has been suggested that in most cases the services that come from nature are valuable enough to make conservation profitable, making money and protecting nature are all too often mutually exclusive goals. Take the case of Africa's Lake Victoria, where the introduction of the invasive Nile perch (*Lates niloticus*) contributed significantly



Locals around Africa's Lake Victoria benefit from trading in the ecologically detrimental Nile perch.

to the decimation of local biodiversity while dramatically boosting the economic value of the lake. Local people profiting from trade in the fish hail its introduction as a success, whereas biologists have condemned the event as "the most catastrophic extinction episode of recent history"⁹. John Terborgh¹⁰, discussing similar issues in tropical-forest conservation, remarked that these forests are "worth more

dead than alive". If Terborgh's assessment is not always true, it is true all too often. So we must directly confront the reality that conservation may be expensive and stop

deceiving ourselves and partners in conservation with hopes that win-win solutions can always be found.

Infinite value

Are there other socially viable paths for conservationists besides the commodification of nature? Yes. Nature has an intrinsic value that makes it priceless, and this is reason enough to protect it. The idea is not new. We view certain historical artefacts and pieces of art as priceless. Nature embodies the same kind of values we cherish in these man-made media. Some ecologists claim that these intrinsic values, often referred to as cultural services, figure prominently enough in their valuation programmes. However, this co-option seems in many cases incongruous. I suggest that the aggregate value of a chunk of nature — its aesthetic beauty, cultural importance and evolutionary significance — is infinite, and thus defies incorporation into any ecosystem service programme that aims to save nature

by approximating its monetary value.

All of this is not to deny a role for ecosystem services in our general efforts to protect nature. Individual ecosystem services will occasionally prove to be useful bargaining chips in specific conservation plans and, as such, can meaningfully support programmes aimed at protecting nature for nature's sake. However, to avoid trading in significant long-term conservation successes for marginal short-term gains, philosophical clarity is essential and caution is needed. When we employ the aid of ecosystem services to help pay the bills of conservation, we must make it abundantly clear that our overall mission is to protect nature, not to make it turn a profit.

Some will argue that this view is simply too optimistic. They may believe that the best way to meaningfully engage policy-makers driven by the financial bottom line is to translate the intrinsic worth of nature into the language of economics. But this is patently untrue — akin to saying that civil-rights advocates would have been more effective if they provided economic justifications for racial integration. Nature conservation must be framed as a moral issue and argued as such to policy-makers, who are just as accustomed to making decisions based on morality as on finances.

The track record of achievements by conservationists motivated by a moral imperative to protect nature for nature's sake is impressive: consider the international ban on commercial whaling, the national parks of the United States, and the CITES ivory-trade ban. Meanwhile, the only 'successful' large-scale ecosystem-service-based conservation project yet achieved is the imperilled Catskill watershed. But this 'nugget' may turn out to be fool's gold.

We will make more progress in the long run by appealing to people's hearts rather than to their wallets. If we oversell the message that ecosystems are important because they provide services, we will have effectively sold out on nature.

Douglas J. McCauley is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

"We will make more progress in the long run by appealing to people's hearts rather than to their wallets."

1. Daily, G. C. *Nature's Services* (Island Press, Washington, DC, 1997).
2. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being: A Framework for the Assessment* (Island Press, Washington, DC, 2003).
3. Hayward, B. *From the Mountain to the Tap: How Land Use and Water Management Can Work for the Rural Poor* (NR International, UK, 2005); available at www.frp.uk.com/assets/Water_book.pdf.
4. Gibbard, S., Caldeira, K., Bala, G., Phillips, T. J. & Wickett, M. *Geophys. Res. Lett.* **32**, L23705 (2005).
5. Woodroffe, R., Thirgood, S. & Rabinowitz, A. (eds) *People and Wildlife, Conflict or Co-existence?* (Cambridge Univ. Press, New York, 2005).
6. Willott, E. *Restor. Ecol.* **12**, 147–153 (2004).
7. Ricketts, T. H., Daily, G. C., Ehrlich, P. R. & Michener, C. D. *Proc. Natl Acad. Sci. USA* **101**, 12579–12582 (2004).
8. DePalma, A. *New York Times* New York's water supply may need filtering (20 July 2006).
9. Wilson, E. O. *The Diversity of Life* (Belknap Press of Harvard Univ. Press, Cambridge, Massachusetts, 1992).
10. Terborgh, J. *Requiem for Nature* (Island Press, Washington DC, 1999).

BOOKS & ARTS

Hitting the right note

Our ability to predict events shapes the way we listen to and appreciate music.

Sweet Anticipation: Music and the Psychology of Expectation

by David Huron

Bradford Books: 2006. 462 pp. \$40

Petr Janata

Imagine that just before sounding the final chord of a Beethoven piano sonata, a concert pianist abruptly interrupts the performance to answer the phone that has just rung in his tuxedo pocket. The audience is left pining for musical closure, livid that a sparkling performance should terminate with an insipid ringtone. It might be forgiven, however, if it was an intentional act of musical humour. In *Sweet Anticipation*, David Huron provides a richly detailed theory of how and why the audience has particular expectations and emotions about such unlikely events, as well as in more common situations.

Scientists trying to understand how we experience music have long assumed that expectations and emotions are intertwined. The basic notion, spurred by music theorist Leonard Meyer's seminal 1956 treatise *Emotion and Meaning in Music*, is that the playful tweaking of a listener's expectations engages general brain mechanisms that compare expectations with input from the environment and set our emotional tone accordingly.

At the core of Huron's account is the premise that organisms derive an evolutionary benefit from being able to predict how their environments will behave. In other words, we internalize the probabilities with which certain events follow other events in everything we do. If I tell my employer I want a raise, there are certain probabilities associated with getting what I want, being ignored or getting fired. My life experiences and understanding of the situation will guide my actions and shape my expectations. Music is no exception. Every time we listen to a blues song or a piano concerto, our brains pick up on the underlying statistics regarding which notes tend to occur together or follow one another in these different styles. We use this accumulated knowledge to appraise unfamiliar pieces of music or different performances of well-known songs. In short, our expectations are an outcome of statistical learning.

Huron does an excellent job of illustrating that there are different varieties of expectation, and that they unfold in time and across different timescales. Expectations are dynamic



Composers often play with our expectations to alter our emotional response to a piece of music.

and shape emotions from one moment to the next. Most of the neuroscientific details of this interplay have yet to be worked out, but Huron provides a compelling framework that is likely to shape research agendas for years to come.

Perhaps the most far-reaching aspects of the book relate to Huron's wedding of contemporary research topics: the cognitive psychology of statistical learning and large-scale statistical analyses of music. The latter topic is important not only for music theorists and researchers, but also for the burgeoning industries of predicting hit songs and customized music-recommendation services. Music researchers face the problem of knowing whether any given piece or set of pieces used in an experiment, from the millions of possibilities, is sufficiently representative to allow broader inferences to be drawn about the psychology and neuroscience of music. Similarly, composers face the problem of deciding whether their compositions play with listeners' expectations in truly original ways.

One solution is to compare the statistics of the piece being considered — how often different notes occur and how often certain pairs of notes or chords follow each other, for example — with statistical distributions from music libraries. This computational-musicology approach allows Huron to answer question after

question about the compositional practices of different composers, genres and cultures, and about the listener's perceptions. The issue of whether feelings such as tension, longing and surprise are associated with less probable musical events is addressed empirically with various examples that simultaneously inform the musical novice about the inner workings of music, and provide the music theoretician with an empirical link between the structure and function of specific compositional practices. As someone who finds Wagner's music difficult to listen to, I realize that my listening experiences would have benefited long ago from reading Huron's concise and insightful treatment of the compositional styles of Wagner, Schoenberg and Stravinsky.

Lest the reader believe that a few improbable examples have been chosen to link statistics, expectation and emotion, Huron offers several variations on the theme. Some, such as explanations of why large melodic intervals (jumps in pitch) tend to go upwards and are followed by smaller steps in the opposite direction, will appeal mainly to music theoreticians and psychologists; others, such as why we grow to like certain pieces of music, why the music of other cultures sounds strange, and the phenomenon of perfect pitch, have broader appeal. Huron's treatment of perfect pitch — the ability to

ARCHIVBERLIN FOTOAGENTUR/ALAMY

name an isolated note without any referent — is particularly refreshing and insightful. This phenomenon is one aspect of the psychology of music that has garnered considerable public and scientific attention. Huron subtly questions the utility of absolute pitch, suggesting that we may have lost this ability over our evolutionary history because the adaptive pressures of our musical environments support listening

to and producing music in terms of relative, rather than absolute, pitch.

Many such intellectual provocations take the reader on a fascinating journey into the inner workings of music and how it tickles the human mind.

Petr Janata is at the Center for Mind and Brain, University of California, Davis, California 95616, USA.

Learning the mother tongue

The Infinite Gift: How Children Learn and Unlearn the Languages of the World

by Charles Yang

Scribner: 2006. 288 pp. \$25

Annette Karmiloff-Smith

Clearly nature and nurture both play vital roles in human development, and most scholars have now rejected the old 'either/or' controversy in favour of theories that invoke complex interactions between the two. However, the debate remains as to whether nature or nurture plays the greater role in shaping the developing brain's acquisition of a native tongue. This is partly because we lack testable theories of the precise processes by which genes and environment interact, and partly because many in the field have entrenched philosophical views about what it is to have human language. For some, consistent regularities in the physical and social environments to which children are exposed play a critical role; for others, the environment simply acts as a trigger for the functioning of our genetic endowment.

In *The Infinite Gift*, Charles Yang navigates between these two positions, endorsing without question Noam Chomsky's 'nativist' theory, which holds that the abstract human capacity for language is prespecified in our genes. But Yang gives the theory a new slant: he argues that children learn their mother tongue by unlearning all the other languages of the world. According to Yang, English children have a different language from adults, not because they occasionally speak imperfect English, but because they occasionally speak perfect Chinese (or Eskimo, French, German, Swahili...). He claims that childhood errors never violate the principles and parameters of the world's languages, and can be explained by translating them into other languages.

Yang argues that, during the early years, children engage in 'variational learning', whereby several languages simultaneously form part of the child's ongoing hypotheses. Gradually, through competition between grammars, the child hones in on the mother tongue: "every instance of language learning

is just a bunch of parameters fighting for supremacy". Inspired by Darwin and using some compelling examples, Yang asserts that language learning and historical change can both be explained by the mechanism of natural selection. Yang's nativist approach seems more dynamic than that of many of Chomsky's disciples, placing more emphasis on learning by the child.

Engagingly written, with fascinating examples conveying the author's enthusiasm for his topic, Yang's book vacillates between targeting parents, students and academic peers. He doesn't mention competing theories, so the naive parent or student would be forgiven for thinking that the whole scientific community concurs on how language is acquired. Yet a growing body of cross-linguistic work based on construction grammar has been emerging from evolutionary anthropology, offering the richest hypothesis-driven data sampling of early child language currently available. In my view, readers of the book would have benefited from being able to contrast chomskyan

theory with at least one other. Instead, we are presented with supporting experimental work emanating solely from Chomsky's disciples. For instance, Yang mentions work on a genetic disorder called Williams syndrome, claiming that moderate to severe mental retardation coexists with unimpaired linguistic ability. But he fails to bring to the reader's attention the work of numerous labs demonstrating that the general intelligence of those with Williams syndrome is actually in the mild range, and that their linguistic ability is seriously delayed in childhood, follows an abnormal developmental trajectory, and reveals semantic, grammatical and pragmatic deficits in adolescence and adulthood.

To support his assertion that there are critical periods for language learning, Yang calls upon examples of feral children. Yet in her book *Wild Boy* (Sceptre, 2003), Jill Dawson made a convincing case regarding the 'wild boy of Aveyron' that it was not the lack of input at appropriate ages that impeded language acquisition, but the fact that the boy was severely autistic (perhaps also explaining why he was abandoned in the first place).

On the topic of genes, Yang claims that "the uniquely human ability for language must ultimately reside in some uniquely human genes". Yet elsewhere he recognizes that *FOXP2*, the so-called 'language gene', is also involved in other cognitive functions and the development of body parts. In fact, it is questionable whether *FOXP2* is directly involved in language at all. Rather, it is indirectly involved in rapid speech perception and production, and is expressed in humans, apes, mice and birds. Despite the allelic difference of *FOXP2* in humans, its function is similar across many species — the rapid coordination of intricately timed motor sequences. The gene is first expressed in many different cerebral regions, but over developmental time its expression is increasingly confined to the cerebellum.

Let me conclude by drawing attention to Yang's justification of his allegiance to a nativist theory; as he says, "we cannot poke around the child's brain". Such an assertion is rather outdated, as we can now examine online processing in infant brains by using high-density evoked related potentials, near-infrared spectroscopy, and other non-invasive methods. These have already shown that the infant cortex starts out highly interconnected, with widespread activity occurring for different inputs. Only with time does the infant brain become specialized and localized for different functions. Such studies are likely to open exciting windows on how the developing brain acquires language.

Annette Karmiloff-Smith is head of the Neurocognitive Development Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.



Speaking a different language: do children start with a grasp of all the world's languages before focusing on their mother tongue?

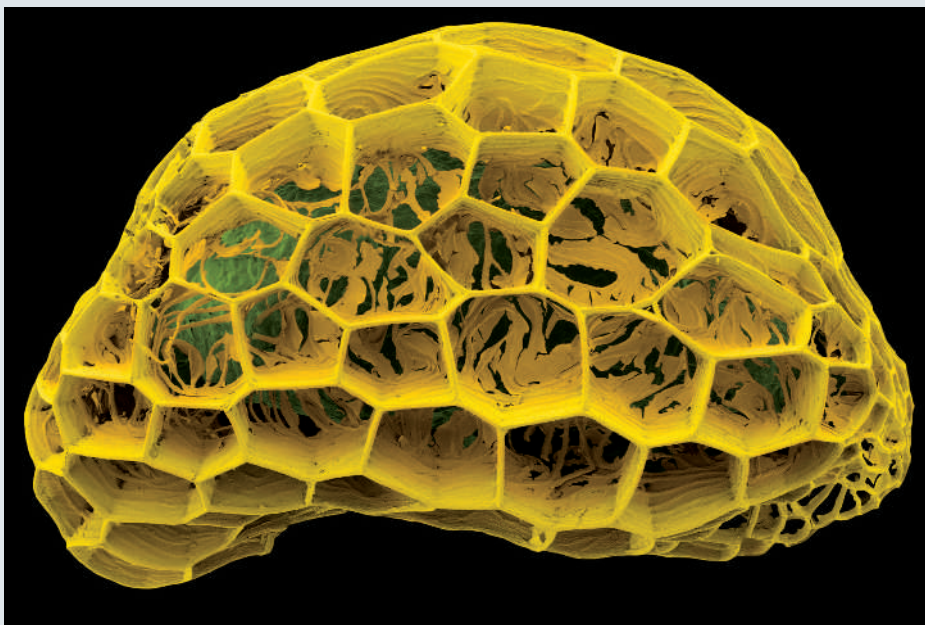
S. WATSON/STONE/GETTY

Blowin' in the wind

Whether clinging to a sock, surfing the breeze or sliding through an animal's gut, seeds have been beautifully designed by evolution for dispersal. They first appeared 360 million years ago and have helped plants colonize much of Earth's surface. True time capsules of life, seeds may travel thousands of miles and wait for many years until they encounter the right conditions for germination.

In *Seeds: Time Capsules of Life* (Papadakis, £35/Firefly, \$60), artist Rob Kessler and seed morphologist Wolfgang Stuppy, both of the Royal Botanic Gardens at Kew, present a natural history of seeds, showcasing their specialized architecture in stunning close-up photographs and scanning electron micrographs. The walls of the 'balloon' seed of the yellow paintbrush (*Castilleja flava*) shown here have dissolved, leaving a honeycombed cage ready to ride the wind.

B.K.



R. KESSLER, W. STUPPY/PAPADAKIS

A pro-darwinian tool-kit

Darwinism and its Discontents

by Michael Ruse

Cambridge University Press: 2006. 328 pp. \$19.99, \$30

Mark Pagel

In the UK House of Lords, peers who vote for a piece of legislation are called the 'contents' and those who vote against it are called the 'not contents'. They are not called 'discontents' or 'malcontents', and this is an important distinction. Being merely not content gives hope that were the legislation to be modified, a not content could move to the content lobby. To be discontented connotes something more extreme — a disaffection or dislike. Worse yet, malcontents evince seditious or even treasonous leanings.

Michael Ruse, then, has probably chosen the title of his new book, *Darwinism and its Discontents*, accurately. What seems clear from the attention received by critics of darwinism — the creationists and the proponents of 'intelligent design' — is that they are more than merely not content with darwinism; they are disaffected, they actively do not like darwinism. The distinction is, again, important.

The book is not a sociological study of the discontents; Ruse does not tell us who they are or even whether they are large in number. Rather, it is a sort of 'battle book', containing facts about evolution and natural selection, and is designed to be instructive in changing people's minds about darwinism. But creationists' disaffection with darwinism may only be cloaked in quibbles about design, the age of Earth, fossils, missing links and the importance of the peppered moth, *Biston betularia*.

These concerns may conceal a deeper affective — indeed a limbic — response to the theory.

This divide between the disaffected and the merely uninformed or misinformed may be why it is so difficult to get anti-darwinists to join the contents. If it is repellent for a creationist to believe facts about evolution and geology, trying to persuade them by using more such unpleasant facts may be a struggle. Daniel Dennett memorably described darwinian theory as a 'universal acid' capable of dissolving all biological facts dropped into it. But religious belief effortlessly bats away empirical and philosophical challenges, making darwinism appear about as acidic as baby lotion. It is as if nothing less than a visit from God himself saying "Don't worry, darwinism is true, I made it that way" will provide the reassurance that believing in evolution is not a one-way ticket to Dante's seventh circle of hell.

Failing help from above, darwinists can only do what they can: indefatigably and hopefully chip away at the edifice of wrong beliefs that underpin the views of discontents. The jacket to *Darwinism and its Discontents* says the book "will appeal" to "concerned citizens who worry that Darwinism is a naturalistic religion that is forced on schoolchildren". It may be optimistic to think it will appeal to these people. But Ruse's biographical chapter on Darwin, an informative survey of how known facts of biology fit the theory of evolution, and well-framed chapters on the origin of life, descent with modification and human evolution, form a concise tool-kit of pro-darwinian chisels. Whether the discontents use these newly acquired chisels on themselves, or have them used on them by

well-meaning friends, teachers or parents, does not really matter.

As a philosopher, Ruse is well placed to discuss darwinism's ethical, religious and philosophical dimensions. If the discontents stay with his book long enough to reach its later chapters, they will find material that transcends the conventional empirical challenges to their disbelief. A common misunderstanding of darwinism is that, somehow, the facts of biology about mere animals are pointers to the characteristics of humankind. This belief has been called the 'naturalistic fallacy', and is a conflation of what is and what ought to be the case, wrongly justifying in humans all sorts of activities from sex discrimination to aggression. More broadly, some believe in what philosophers call 'evolutionary epistemology' — that the ways we behave and the views we hold must be accurate reflections of the true nature of the world, otherwise natural selection would not have favoured them.

Of course, deductions from evolutionary epistemology, although undoubtedly sometimes true, are formally wrong, not least because culture has allowed humans, probably uniquely in degree, to weaken the links to their biology. Ruse sums it up nicely: "No one is surprised that Mendel, who was right about heredity, had no children, and that Darwin, who was wrong about heredity, had ten children." Equally, this seemingly obvious point should give the discontents of Ruse's title pause. Religion is found in all human cultures, and most include creation stories not unlike those of the major religions. But this universality, although perhaps signalling some human predilection to hold religious views, should not necessarily reassure: superstition is also universal. ■ Mark Pagel is professor of evolutionary biology, School of Biological Sciences, University of Reading, Reading RG6 6AJ, UK.

Reflections on Wallace

An unpublished paper has recently come to light, which shows that even at an early age, Alfred Russel Wallace was bold enough to approach the scientific establishment with his ideas.

Charles H. Smith

The image of Alfred Russel Wallace as a self-effacing satellite to Darwin's more resolute brilliance has taken a fresh blow. An early, unpublished Wallace paper illustrates his boldness in approaching leading scientists with his ideas. In the spring of 1843, some 15 years before he wrote to Darwin setting out his thoughts on evolution, Wallace described possible ways to improve the mirrors used in telescopes in a letter and paper sent to William Henry Fox Talbot. At that time, Talbot was a leading man of science, already famous for his invention of a successful early form of photography.

I chanced upon this paper earlier this year as the result of a Google search. Finding that an 1843 letter by "Alfred R. Wallace, surveyor" had been transcribed online as part of a project called "The Correspondence of William Henry Fox Talbot", I contacted the project's director, Larry Schaaf, and asked whether the paper mentioned in the letter still existed. It did. Even the envelope in which the letter and manuscript were originally delivered to Talbot had been preserved. The paper is reproduced on the next page.

Having been forced to leave school at 13, Wallace had been working for six years for his older brother William, as a surveyor at various locations in the west of England and South Wales. But the curators of the Fox Talbot correspondence project had not realized this young surveyor was the Wallace who is now famous as one of the founders of the theory of natural selection.

At this point Wallace had not met his future colleague Henry Walter Bates, corresponded with or met Darwin, read the anti-creationist best-seller *Vestiges of the Natural History of Creation* (it was published the following year) or shown any real interest in the problem of the origin of species, which would lead to his most famous work.

Schaaf is a renowned authority on the early history of photography. He says the ideas Wallace expressed in his paper addressed challenges in optical engineering that may have been too difficult technologically to implement at the time.

According to Schaaf, Wallace was basically saying that gravity pulls a pool of mercury into a near-perfect plane and that by electroplating another metal on to the layer of mercury, one could form a perfectly flat mirror.

Later on in the paper, Wallace discusses curved mirrors. Schaaf says that at the

time, these were cast in alloys and then ground by hand to the best trial and error curvature possible. They tarnished quickly and had to be constantly repolished, without changing the curvature. "Astronomers spent as much time doing that as they did looking at the heavens," explains Schaaf. "So it was a real problem and Wallace and Talbot were exploring different but related approaches to solving it." Talbot was obviously interested enough to keep the letter and paper, although it is not known whether he ever sent Wallace a reply.

But this may not be the end of the story. A technology known as 'liquid spinning mirror telescope' is also based on the use of mercury, and is currently being developed at several locations around the world. According to Paul Hickson, the leader of one group that's developing it at the University of British Columbia in Vancouver, the first mention of the liquid-mirror technique can be found in a published letter by the astronomer Ernesto Capocci.

Capocci was a colleague of and corresponded with the noted optician and microscopist Giovanni Battista Amici, who in turn had received Talbot in Italy as a guest several times since the early 1820s and exchanged at least 24 letters with him. In fact, Amici visited Talbot while in London in 1844 on optics-related business, and less than three years later, his son Vincenzo, a mathematician, published a paper entitled "Considerazioni sulla teoria del moto dei liquidi" (Considerations on the theory of motion in liquids). So it is possible that some of Wallace's ideas found their way into published work after all.

Wallace was just 20 years old when he wrote the paper on mirrors, and it is, by seven years, his earliest extended writing on a technical subject that we know of. His first published paper, a short characterization of the Amazonian umbrella bird, appeared in 1850.

It is interesting that the young and self-educated Wallace had enough confidence in his idea to set it before the leading English thinker of the time on such matters. Fifteen years later he would do the same with Darwin. In his autobiography, Wallace describes how he never shied away from debate once he felt he had firmly grasped the basic elements of a question. When it came to serious discussion, mere weight of reputation meant little to him.

His correspondence with Talbot also helps us appreciate certain threads in Wallace's intellectual life that are often



NATURAL HISTORY MUSEUM LONDON

Intrepid: Alfred Russel Wallace did not shy away from debate with leading scientists.

peripheralized in discussions of his contributions to evolutionary biology and biogeography. Remarks in his autobiography and several secondary sources suggest that he had mastered basic principles of optics, surveying, geodesy and astronomy by the age of 18 or 20.

This 'quickness' would later support his entry into a variety of debates. And Wallace's strong interest in optics and photography would inform several of his more unusual investigations, such as his interpretation of the canal-like structures appearing on contemporary photographs of Mars, over which he argued with the astronomer Percival Lowell.

My discovery of this paper underlines the increasing value of the Internet as a means of identifying archival sources relevant to ongoing research. Were it not for Schaaf's efforts in making the correspondence of Fox Talbot electronically accessible, this particular item might have remained undiscovered for another 163 years. ■

Charles H. Smith is Professor of Library Public Services and Science Librarian at Western Kentucky University, Bowling Green, Kentucky, USA.

FURTHER READING

Fichman, M. *An Elusive Victorian: The Evolution of Alfred Russel Wallace* (Univ. Chicago Press, Chicago, 2003).
The Correspondence of William Henry Fox Talbot <http://foxtalbot.dmu.ac.uk/>
The Alfred Russel Wallace Page <http://www.wku.edu/~smithch/index1.htm>
The British Library has recently obtained the original Wallace letter and paper; transcriptions of each may be viewed online at <http://foxtalbot.dmu.ac.uk/letters/transcriptDocnum.php?docnum=4807>

ESSAY

On a probable means of procuring plane and curved Specula of Great size, with a few remarks on fixed telescopes. by A. R. Wallace.

On considering the means of obtaining in the first place a perfect plane Reflecting surface, it appears that nature in certain instances produces one much superior to any thing the art of man can form. The surface of quiescent fluid mercury is perhaps the best of these and all we want is to fix it immoveably or obtain a perfect copy of it. There are many difficulties in the way of the latter but it is not unlikely that the Electrotypes offers a means of overcoming them, and producing in solid metal a truly plane polished mirror. There appears no reason why the operation should not be successful; the vessel containing the metallic solution might have a stratum of mercury at the bottom and the moment the connection was made by the wire touching the mercury the operation would commence, an almost infinitely thin pellicle being first deposited the surface must be accurately copied, and in fact it has been shewn that a surface equally polished with the copy is produced. The difficulty of keeping the vessel perfectly free from vibration might doubtless be got over; the only question then remaining is a Chemical one: what metal can be deposited that would not in any degree combine with the mercury? Neither Gold nor Copper would do, perhaps Iron or Antimony might do better, but that which would answer must be determined by experiment. The writer much regrets that the circumstances in which he is placed have not permitted him either the time or expense necessary to ascertain these points. Should the metal deposited not reflect sufficient light he conceives a great advantage would still be obtained over performing the whole operation by hand as a thin pellicle of Gold or other metal might be deposited over it and that if necessary polished. Even should all efforts to deposit a metal on Mercury so as to produce a polished surface, fail, the principle might still perhaps be applied by obtaining some fluid which would consolidate with a polished surface. it might then be metallized in the manner usually employed for Electrotypes operations and though perhaps a highly polished surface might not be thus obtained still it would be a great assistance to have a true homogenous plane surface requiring nothing but the last polish to render it fit for use. Trusting however that whatever difficulties may be presented will be overcome he will proceed to point out the advantages that may result from its success. We should then have a perfect plane mirror of any size to reflect the rays to the concave speculum and thus one great difficulty of using telescopes of large size fixed, would be got over.

Having however obtained a plane mirror it appears not impossible that we may also produce the concave one without going through the tedious, uncertain, and expensive operations of grinding and polishing. It is evident that a plane mirror produced as before described would be both perfectly uniform in its thickness and perfectly homogenous in its structure. The writer of this was much interested some time ago with Mr. Nasmyth's method of producing concave reflectors by extracting the air from a chamber behind a circular plate of glass and allowing the pressure of the atmosphere to produce the required curvature. He has not been able to learn why this has not succeeded or been brought into operation, but supposes it must be from the difficulty of procuring plates of glass sufficiently uniform in thickness, homogenous in structure, and truly plane in both surfaces without all of which it is evident a uniform curve would not be produced. Mr. Nasmyth conceives that the curve would be between a catenary and circle, but does it not appear more likely that it would be truly parabolic, the pressure of the atmosphere being similar to an infinite number of perpendicular pressures. If so it appears difficult to imagine how a perfectly plane homogenous body of uniform thickness, placed under such a pressure could take any other than a true parabolic form. If it did there would be a curved mirror formed superior probably to any yet made. It also seems probable that the metal after being exposed to the pressure, would on its being removed retain a portion of its curvature sufficient for a long focus and thus obviate the inconvenience of having to keep the atmospheric pressure perfectly uniform in order to preserve the same focal length.

In fixing a telescope for the purpose of using a moveable plane Speculum to reflect the objects to the curved one, it appears evident that that position must be best which requires the plane mirror to be of the least possible size. To effect this the tube must be directed to a point midway between the Pole and a few degrees above the Southern Horizon or in this Country about 20° South of the Zenith, (on the supposition that it is not required to view objects more than a few degrees below the Pole) the Great curved Speculum being at the top and the plane mirror and observer at the bottom; In this situation the plane mirror need be only 1/7th larger than the Speculum to reflect rays to the whole of its surface from objects least favourably situated, while if directed toward the Pole with the Speculum at the top the plane mirror must be nearly twice the diameter and with the speculum at bottom, no dimensions of plane mirror would reflect light from objects near the Pole to the whole surface of the Speculum.

The plane mirror might be made to keep an object in the field of view with great facility in whatever direction the telescope were placed, by fixing it in the manner of an equatorial so as to be elevated to the proper angle and moved round on a polar axis.

The writer has thus concisely expressed his ideas on these subjects in the hope that if they are at all novel, some one may be found to put them to the test of experiment and ascertain whether these methods of increasing the power of the Telescope are capable of practical application.

CANCER BIOLOGY

Infectious tumour cells

David Dingli and Martin A. Nowak

Cancer cells are generally viewed as a problem innate to their host, but evidence is mounting that they can evolve to become infectious agents and be transmitted between individuals.

The current view of cancer development is that normal cells are transformed into tumour cells by sequential mutations that activate cancer-promoting 'oncogenes', or inhibit genes that would otherwise suppress tumours, or trigger genetic instabilities. As a consequence, every tumour is the result of a unique evolutionary process as the cancer cells adapt to out-compete their neighbours. This process begins in its individual host and ends with the elimination of the tumour or the death of the host. Two studies^{1,2}, however, suggest that some tumour cells can behave like infectious agents and move from one host to another.

It has long been suspected that canine transmissible venereal tumour (CTVT) is transferred between dogs by implantation of tumour cells from the donor to the recipient, where the tumour grows as a graft (Fig. 1)³. Several lines of evidence provided indirect support for this hypothesis. A tumour can only be induced by the implantation of whole CTVT cells, and not by cell extracts or dead cells. Normal canine cells have 78 chromosomes, but the tumour cells isolated from different animals characteristically have 58 to 59 chromosomes⁴. In addition, a particular insertion near the *c-myc* oncogene is present in all tumour samples⁵. If the graft transfer hypothesis is correct, tumour cells from different animals should all be similar genetically, and should be different from the normal cells of the host animal.

Writing in *Cell*, Murgia *et al.*¹ now provide a formal proof of the graft hypothesis for CTVT. They studied CTVT cells and normal cells isolated from dogs from five continents. They show that all the tumours are closely related genetically, and different from normal cells of the host dog, by using a combination of genetic techniques (dog leukocyte antigen (DLA) haplotyping, microsatellite DNA and mitochondrial DNA sequencing).

The tumours from different animals had less genetic variability than that observed within even the most inbred breed of dog. Therefore, the tumours could not have arisen from the separate cancerous transformations of cells within individual animals, but have been transmitted from one dog to another, spreading from an ancestral clone. Murgia *et al.*¹ estimate that

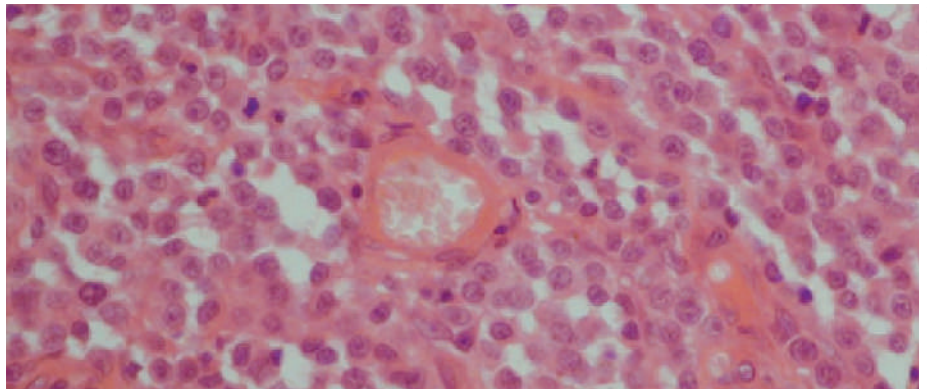


Figure 1 | A canine transmissible venereal tumour in the progressive stage.

this clone arose between 250 and 2,500 years ago, making it the oldest known continuously replicating cancer cell line in animals¹.

In an independent study, Pearce and Swift² reported that devil facial tumour disease is caused by tumour cell transmission between Tasmanian devils. Cancer cells isolated from different animals with tumours of different age and size have the same pattern of chromosomal abnormalities. Furthermore, all the normal cells from one animal had part of one chromosome inverted, but this inversion was not found in any of the tumour cells from that animal, suggesting that the tumour did not arise from host cells.

So, can infectious transmission of cancer also occur between people? There is no evidence (yet) for direct person-to-person transmission of tumour cells during normal social contact. The only known physiological route for tumour cell transmission in humans is during pregnancy. In the United States every year, about 3,500 pregnant women also have a malignancy, and transplacental transmission of acute leukaemia, lymphoma, melanoma and carcinoma from mother to fetus has been observed⁶. Acute leukaemia cells have also been transferred between fetuses in mothers with a multiple pregnancy, followed by the development of disease in both fetuses⁶.

Organ transplantation is another possible route of tumour cell transmission between people. The immunosuppressive therapy required for survival of the transplanted organ blunts

the immune surveillance that might otherwise recognize and react against donor-derived tumour cells. Fortunately, the development of donor-derived tumours is rare: 0.04% of solid organ transplant recipients develop a cancer that was transferred from their donor⁷, and 0.06% of recipients of haematopoietic stem-cell transplants develop blood cancers from the transferred cells⁸. The main culprit seems to be malignant melanoma that is undetected in the donor at the time of organ harvest⁷. Finally, we came across a case report where a surgeon developed malignant fibrous histiocytoma after accidentally injuring his palm during surgical removal of the tumour from a patient⁹.

Why is cancer in general not transmissible between people? A main reason is tissue graft rejection caused by so-called MHC incompatibility. In humans and other vertebrates, the immune system uses cell-surface proteins called MHC antigens to differentiate 'self' from 'non-self' cells because these proteins vary between individuals. The immune system then reacts against those cells that display non-self MHC antigens. Such reactions may protect against tumour cell engraftment by eliminating implanted cells.

In support of this idea, CTVT cells avoid immune-mediated destruction by down-regulating the expression of the canine MHC antigens (the DLAs)¹. This adaptation of the cancer cells is crucial because complete absence of DLAs would allow an immune response in which natural killer cells destroy the tumour,

C. MURGIA

whereas normal expression of DLAs activates a different immune reaction involving killer T cells with a similar outcome. In many dogs, an immune attack against CTVT ultimately develops, leading to eradication of the tumour and immunity to reinfection³. This evidence also strengthens the hope that the immune system might be coaxed into eradicating established tumours in humans.

The emergence of multicellular life forms required cooperation between the cells of an individual organism. Cancer entails loss of this cooperation, and from the perspective of evolutionary game theory, cancer is a 'defector'¹⁰. Breakdown of cooperation can lead to the death of the host, but the tumour also meets its own demise. Therefore, a tumour that can be transmitted from one host to the next manoeuvres around the specific evolutionary mechanism that is meant to control it.

MHC incompatibility means that the cancer cells of a donor should induce a vigorous immune response in a healthy recipient. Indeed, tumour transfer in mice is only possible between animals that share the same MHC, or if the recipient is severely immunosuppressed. This leads to the interesting speculation, suggested by Murgia *et al.*¹, that a main

reason for MHC diversity in humans and other vertebrates is to ensure that cancer is not normally an infectious disease. If so, it seems that certain cancers have evolved mechanisms to get around these safeguards.

David Dingli and Martin A. Nowak are in the Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA. David Dingli is also at the Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. e-mails: dingli.david@mayo.edu; martin_nowak@harvard.edu

1. Murgia, C., Pritchard, J. K., Kim, S. Y., Fassati, A. & Weiss, R. A. *Cell* **126**, 477–487 (2006).
2. Pearce, A. M. & Swift, K. *Nature* **439**, 549 (2006).
3. Das, U. & Das, A. K. *Vet. Res. Commun.* **24**, 545–556 (2000).
4. Wright, D. H., Peel, S., Cooper, E. H. & Hughes, D. T. *Rev. Eur. Stud. Clin. Biol.* **15**, 155–160 (1970).
5. Katzir, N., Arman, E., Cohen, D., Givol, D. & Rechavi, G. *Oncogene* **1**, 445–448 (1987).
6. Tolar, J. & Neglia, J. P. *J. Pediatr. Hematol. Oncol.* **25**, 430–434 (2003).
7. Kauffman, H. M. *et al. Transplantation* **74**, 358–362 (2002).
8. Sala-Torra, O. *et al. Biol. Blood Marrow Transplant.* **12**, 511–517 (2006).
9. Gartner, H. V. *et al. New Engl. J. Med.* **335**, 1494–1496 (1996).
10. Nowak, M. A. in *Evolutionary Dynamics: Exploring the Equations of Life* (Belknap/Harvard Univ. Press, Cambridge, MA, 2006).

step beyond chaos, but it forms a well-defined, persistent steady state.

Hof *et al.*¹ cast serious doubt on this thesis. Although they focus on flow in a straight pipe (Fig. 1), their results may well apply to the broader class of turbulent 'shear flows', which are defined by different layers of a fluid moving at different speeds. Such flows include those over cars and aircraft wings and within air ducts, and those used to cool electronics. The authors' observations might also — more controversially — apply to certain flows in astrophysics. What all these cases have in common is that it is unclear when turbulence ensues, and why it persists.

The authors argue, surprisingly, that turbulence started by a gentle kick will eventually decay, no matter how large the nonlinearities in play. A quantity known as the Reynolds number is used to measure the strength of nonlinearity in a flow relative to the smoothing effects of viscosity. Past studies had indicated that shear turbulence could decay at small Reynolds numbers, but that beyond a certain Reynolds number, the turbulence would become permanent^{2,3}. Not so, say Hof and colleagues: shear turbulence is always destined to return to a laminar state — provided, that is, one waits long enough.

The catch is that the wait may be very long indeed. For typical flows, say in a city water main of 60-centimetre diameter, extrapolating the authors' result gives a decay time of 10^{3,000} years! This trapping in a transient state is in many ways comparable to the situation in glasses, which are kept from their equilibrium crystalline state by a painfully slow decay. One can speculate that this new insight might allow controls to be engineered such that, using only very small nudges and the theory of chaos control⁴, turbulent flows might be made to decay to a laminar state sooner.

By comparison, the implications for astrophysical states are less certain, but potentially grander. Shear flows occur in accreting systems, such as protoplanetary systems and around compact objects in galactic cores, such

FLUID DYNAMICS

Turbulence lost in transience

Daniel Perry Lathrop

Turbulence is generally regarded as a permanent feature of many fluid flows. That assumption is challenged by the claim that shear turbulence has a limited lifetime — albeit sometimes a very long one.

Turbulence is not a trivial problem. It influences all manner of physical effects — aerodynamic drag, cooling and heating, to name just a few. And perhaps its most notorious guise is the worrisome mid-flight buffeting that can so upset air passengers. Yet, for all its ubiquity, scientists, engineers and mathematicians have long struggled to understand the underlying nature of turbulence. On page 59 of this issue, Hof *et al.*¹ stir things up further with the suggestion that, far from being the steady-state phenomenon it was assumed to be, turbulence does, in fact, disappear — given time.

Although fluid turbulence most often occurs because a fluid's nonlinearity is great enough to overcome the effects of viscosity, roughness of surfaces and external shaking can play their part, too: with very smooth walls and in a quiet environment, regular, laminar flows can sometimes persist, even at very high flow rates. Many studies, including the Hof paper¹, focus on this ideal case.

Part of the problem in dealing with turbulence in any more practical situation is that its irregular patterns bear little resemblance to the

smooth analytical solutions that emerge from the equations of motion that govern fluid flow. For decades, the general view has been that turbulence is merely a nonlinear solution of these equations: turbulence might be just one

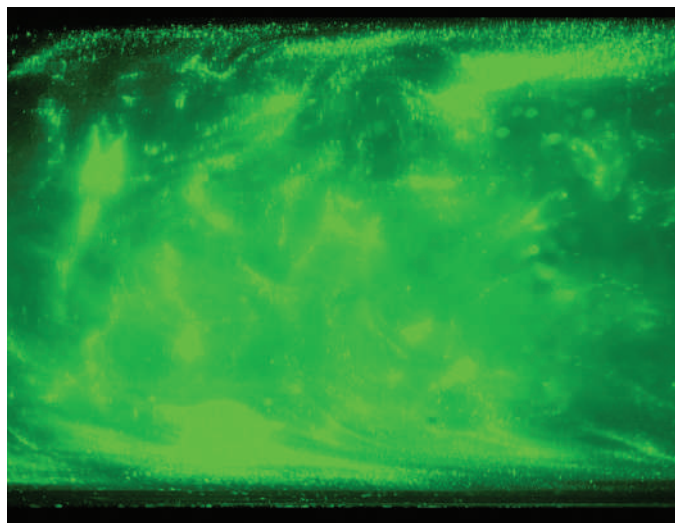


Figure 1 | Inner turbulence. Here, turbulence in a pipe at a Reynolds number of 5,000 is visualized using microscopic crystalline platelets illuminated with a sheet of laser light. These platelets align with shear flows, and changes seen across the flow indicate turbulent fluctuations. Hof *et al.*¹ argue that turbulent flows in pipes such as these are a transient phenomenon.

G. P. BEWLEY & D. P. LATHROP

as black holes⁵. Typically, Reynolds numbers in astrophysics are much larger than the values of a few thousand dealt with by Hof *et al.*¹, indicating a much greater propensity for turbulent flow. If any turbulence in these astrophysical flows were transient over long timescales, this might explain the enhanced accretion and in-fall rates frequently observed, without the need to invoke additional instabilities. Such instabilities are generally postulated to be caused by magnetic fields^{6,7}. Because of the significant interaction of rotation and turbulence in astrophysical objects, however, we cannot be sure that Hof and colleagues' results¹ would apply.

From the applied mathematician's perspective, the implications of these investigations are fascinating. The motions are best understood by thinking about a 'phase space' of configurations of velocities present in a fluid. In this depiction, a system evolves in time by moving around the phase space. The laminar state has a constant velocity field that is smooth and steady in time. In velocity phase space, this is represented by a single point, with trajectories pointing inward from every direction indicating its stability.

For turbulent flow, however, a great tangle of trajectories would show up some distance from the laminar point. The turbulent state amounts to wandering on that tangle, and getting lost on it for considerable periods of time. If Hof *et al.* are correct, however, the system always — however long it takes — finds a connecting trajectory that causes it to fall back onto the laminar attracting point. The exponentially long times indicate a change in the character

of the turbulent tangle that begs for a deeper understanding.

The results¹ do need to be verified by other means and in other shear geometries. Previous studies of the turbulent–laminar transition^{2,3}, including a study from one of the paper's authors², indicated that the turbulent lifetime becomes infinite at Reynolds numbers above around 2,200. This stark contrast between previous and the present results will surely generate considerable activity. The biggest difficulty here will be that the long lifetimes involved limit the range of Reynolds numbers that are practicably testable.

Understanding the idealization of very smooth walls and a quiet environment will also require further study. Beyond that, the possibility that these results could be extended to other shear flows, and the potential for new turbulence controls, will continue to generate excitement in this area of research.

Daniel Perry Lathrop is in the Departments of Physics and Geology, and at the Institutes for Physical Sciences and Technology, and for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 20742, USA.

e-mail: lathrop@umd.edu

- Hof, B., Westerweel, J., Schneider, T. M. & Eckhardt, B. *Nature* **443**, 59–62 (2006).
- Faisst, H. & Eckhardt, B. *J. Fluid Mech.* **504**, 343–352 (2004).
- Peixinho, J. & Mullin, T. *Phys. Rev. Lett.* **96**, 094501 (2006).
- Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **64**, 1196–1199 (1990).
- Balbus, S. A. & Hawley, J. F. *Rev. Mod. Phys.* **70**, 1–53 (1998).
- Velikhov, E. P. *J. Exp. Theor. Phys.* **9**, 995 (1959).
- Sisan, D. R. *et al. Phys. Rev. Lett.* **93**, 114502 (2004).

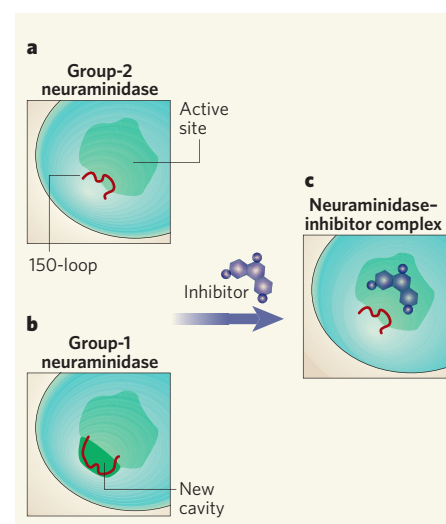


Figure 1 | Induced fit of an enzyme inhibitor. The influenza neuraminidase enzyme is blocked by antiviral drugs. Subtypes of the enzyme exist and are divided into two groups, which differ in the conformation of their active sites. **a**, A loop of amino acids (known as the 150-loop) in the active sites of group-2 enzyme subtypes is arranged such that no conformational changes occur in the active site on binding of an inhibitor. **b**, For group-1 subtypes, the 150-loop has a different conformation from group-2 subtypes, exposing a large cavity next to it. **c**, When an inhibitor binds to group-1 subtypes, the 150-loop adopts a conformation similar to that of group-2 neuraminidases.

complexes, showed that the three-dimensional structure of the enzyme active site is the same for all of these subtypes, and that the inhibitors bind to the active sites in the same way.

The inhibitors don't just suppress the activity of the enzyme subtypes that were used to direct drug development — they have similar activity against N1 neuraminidase, and are effective against flu viruses that contain an N1 neuraminidase⁶. It was always assumed that the inhibitors bind to the active sites of the group-1 neuraminidases in the same way as for group-2 enzymes, as the amino-acid sequences of the active sites are essentially the same for all the subtypes.

However, the crystal structures of the N1, N4 and N8 neuraminidases, reported by Russell *et al.*¹, surprisingly reveal that the active sites of these group-1 enzymes have a very different three-dimensional structure from that of group-2 enzymes. The differences lie in a loop of amino acids known as the 150-loop, which has an unexpected conformation that opens up an adjacent cavity. The 150-loop contains an amino acid designated Asp 151; the side chain of this amino acid has a carboxylic acid that, in group-1 enzymes, points away from the active site as a result of the 'open' conformation of the 150-loop (Fig. 1). The side chain of another active-site amino acid, Glu 119, also has a different conformation in group-1 enzymes compared with group-2 enzymes.

The Asp 151 and Glu 119 amino-acid side

STRUCTURAL BIOLOGY

Antiviral drugs fit for a purpose

Ming Luo

Did drug researchers have a lucky break when they developed antiviral drugs for influenza? Crystal structures of enzymes from the H5N1 virus suggest that they did, and provide avenues for further exploration.

Oseltamivir (Tamiflu) and zanamivir (Relenza) have been stockpiled by several nations to counter the threat of a flu pandemic, should the highly pathogenic avian influenza virus H5N1 develop into a human strain. These drugs are inhibitors of an enzyme known as neuraminidase, which is found on the surface of the flu virus. But would these medications hold their own against a pandemic? Some answers might be found from the crystal structure of H5N1 neuraminidases bound to the antiviral drugs, as reported by Russell *et al.*¹ on page 45 of this issue*. These structures also reveal a cavity in the active site of certain neuraminidase subtypes that could be exploited to

make more effective antiviral drugs.

Avian flu belongs to the genus of influenza virus known as type A, which is divided into nine subtypes based on the variety of neuraminidase expressed. The subtypes fall into two groups²: group-1 contains the subtypes N1, N4, N5 and N8, whereas group-2 contains the subtypes N2, N3, N6, N7 and N9. The enzyme facilitates the spread of virus during an infection; oseltamivir and zanamivir strongly inhibit neuraminidase activity, so limiting the disease. These inhibitors were originally developed using crystal structures of neuraminidase subtypes N9 and N2 and another neuraminidase from the type B genus of influenza viruses^{3–5}. The crystal structures, along with those of neuraminidase–inhibitor

*This article and the paper concerned¹ were published online on 16 August 2006.

chains form critical interactions with neuraminidase inhibitors. For neuraminidase subtypes with the open conformation of the 150-loop, the side chains of these amino acids might not have the precise alignment required to bind inhibitors tightly. If so, influenza viruses that contain these subtypes would be resistant to oseltamivir and zanamivir. However, the crystal structure of N1 neuraminidase in a complex with oseltamivir shows that the open polypeptide loop adopts the conformation observed in the N2 neuraminidase-inhibitor complex. In other words, when an inhibitor binds, the active site adopts the conformation seen in the crystal structures of group-2 enzymes³⁻⁵. This 'induced fit' restores the contacts between the amino-acid side chains of the N1 active site and the inhibitor, which probably explains why the inhibitors are effective against N1 influenza viruses.

The difference in the active-site conformation between the two groups of neuraminidases must be caused by changes to amino acids that lie outside the active site. This means that an enzyme inhibitor for one target will not necessarily have the same activity against another with the same active-site amino acids and the

same overall three-dimensional structure. We are fortunate that the group-1 influenza virus neuraminidases assume a tight interaction with their inhibitors. Otherwise, these antiviral drugs might be ineffective against influenza viruses such as H5N1.

Russell and colleagues' discovery¹ cautions us to pay more attention to changes in amino-acid sequences outside the active site of neuraminidases. Such changes might result in a mutant virus that is resistant to current drugs if the conformation of the active site is altered enough to prevent an induced fit of the inhibitors. We must therefore determine the binding strength of our neuraminidase inhibitors for each new strain of influenza virus, even if the amino-acid sequence of the neuraminidase active site is unchanged compared with that of existing viruses.

More positively, the cavity near the active site that is exposed by the open conformation of the 150-loop might be exploited by drug designers. The affinity of enzyme inhibitors depends on their fit in their respective active sites. One could design a chemical group with the right shape, size and electronic charge to fit snugly

into the newly discovered cavity. Attaching this group to the scaffolding of existing inhibitors could then produce a class of neuraminidase inhibitor that is more effective against H5N1-like flu viruses.

When combating pathogenic microorganisms, either bacteria or viruses, we often face an enemy that continually alters its tactics. The flu virus is no exception. The way to stay ahead in this game is to constantly track the changes of the pathogen and adapt countermeasures accordingly. The crystal structures reported by Russell *et al.*¹ provide valuable intelligence in the war against influenza. ■

Ming Luo is in the Department of Microbiology, University of Alabama, Birmingham, 1025 18th Street South, Birmingham, Alabama 35294, USA. e-mail: mingluo@uab.edu

1. Russell, R. J. *et al.* *Nature* **443**, 45–49 (2006).
2. Thompson, J. D., Higgins, D. G. & Gibson, T. J. *Comput. Appl. Biosci.* **10**, 19–29 (1994).
3. Kim, C. U. *et al.* *J. Am. Chem. Soc.* **119**, 681–690 (1997).
4. von Itzstein, M. *et al.* *Nature* **363**, 418–423 (1993).
5. Taylor, N. R. *et al.* *J. Med. Chem.* **41**, 798–807 (1998).
6. Warren, M. K. *et al.* *Antimicrob. Agents Chemother.* **46**, 1014–1021 (2002).

MATHEMATICS

Controlling our errors

Barry Mazur

The Sato-Tate conjecture holds that the error term occurring in many major problems in number theory conforms to a specific probability distribution. That conjecture has now been proved for a large group of cases.

Even under the best circumstances, controlling our errors is a dicey business. But as a recent series of mathematical papers¹⁻³ shows, significant strides are being made — in number theory at least. The new work amounts to a proof of a 40-year-old conjecture, known as the Sato-Tate conjecture, for a class of mathematical problems with applications in cryptography and the high-speed factorization of large numbers. This conjecture predicts the probability distribution of the error terms that pop up in these problems.

In any empirical study, errors accumulate for many reasons. All an experimentalist can hope for is to know the sources of most errors, and to be able to estimate how much trouble they cause.

The web page of the US Bureau of Transportation, for example, lists six possible causes of systematic error in the 1993 US Census count: inability to obtain information about all cases in the sample; response errors; definitional difficulties; differences in the

interpretation of questions; mistakes in coding or recoding the data obtained; and other errors of collection, response, coverage and estimation⁴.

In pure mathematics, however, whenever we calculate what we hope is a good approxima-

tion to a quantity describing a phenomenon, we also have a shot at understanding — in depth, and beyond a shadow of a doubt — the nature of the ensuing error term, defined simply as: error term = exact value — our 'good approximation'.

Of course, if our approximation is at all good, the error term should be small. For many problems, the gold standard of goodness is what we might call square-root accuracy: that the error term scales as the square root of the quantity as it, and the phenomenon it describes, grows larger and larger. Great successes in controlling error terms in number theory were achieved in the past century. Specifically, through the work of Helmut Hasse⁵ in the 1930s, André Weil⁶ in the 1940s and Pierre Deligne⁷ in the 1970s, a large class of major approximations were proved to have this kind of accuracy.

Take a (randomly chosen) example. For prime numbers p , define $N(p)$ as the number of ways in which p can be written as a sum of 24 squares of whole numbers. (Squares of positive numbers, negative numbers and zero are all allowed.) The ordering of the squares of the numbers that occur in this summation also counts. Thus, the first prime number, 2, can already be written as a sum of 24 squares of whole numbers in 1,104 ways, because there are that many different ways in which two choices of either $(+1)^2$ or $(-1)^2$ can be arranged in a line where the 22 other numbers are zeros.

We know, then, that $N(2) = 1,104$. What about $N(p)$ for the other prime numbers $p = 3, 5, 7, 11, \dots$? A good

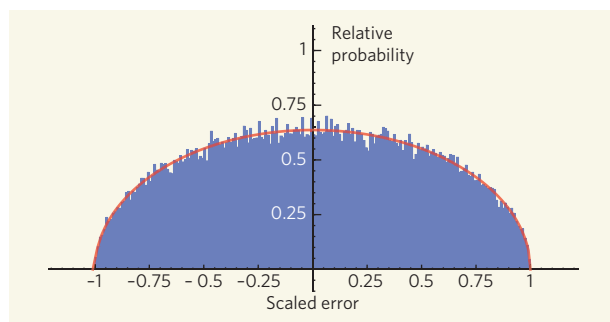


Figure 1 | Probability distribution of error terms. The Sato-Tate distribution $\frac{2}{\pi} \sqrt{1-x^2}$, the smooth red curve in this figure, can be compared with the probability distribution of scaled error terms (blue bars) for the number of ways $N(p)$ in which a prime number p can be written as a sum of 24 square numbers. The data, tabulated for primes p less than a million, agree closely with the distribution, and give hope that the Sato-Tate conjecture holds for this problem. (Courtesy of W. Stein; for details of the computation, see ref. 10.)

approximation for these values $N(p)$ turns out to be $\frac{16}{691}(p^{11} + 1)$. The error term on this approximation, defined as

$$\text{Error}(p) = N(p) - \frac{16}{691}(p^{11} + 1),$$

has been proven to be square-root small⁷. As a matter of fact, $\text{Error}(p)$ is no larger in absolute value than $\frac{66304}{691}\sqrt{p^{11}}$.

This type of precision in obtaining 'good approximations' to difficult mathematical problems and estimating their error has led mathematicians to consider the next, but significantly deeper, tier of questions: determining the probability distribution of those error terms whose magnitudes have been shown to be square-root small.

For example, in the problem described above, let us rescale our error terms by their proven maximum order of magnitude, and ask for the probability distribution of the numbers

$$\frac{\text{Error}(p)}{\frac{66304}{691}\sqrt{p^{11}}} = \frac{N(p) - \frac{16}{691}(p^{11} + 1)}{\frac{66304}{691}\sqrt{p^{11}}}$$

as p ranges through the primes. These numbers all lie in the interval between -1 and $+1$. In 1960, Mikio Sato (by studying numerical data) and John Tate (following a theoretical investigation)⁸ predicted that the absolute values of the scaled error terms for data in many problems of current interest conform to a specific probability distribution; Sato and Tate shared the Wolf prize in 2003 (ref. 9). In this particular instance, this is the simple distribution curve $\frac{2}{\pi}\sqrt{1-x^2}$, where x ranges through the interval between -1 and $+1$ (Fig. 1). Although the Sato–Tate prediction remains unproved for this specific example, the agreement with numerical data for the first million prime numbers gives cause for optimism.

In March this year, extraordinary strides were made towards demonstrating the truth of the Sato–Tate conjecture for one class of problems^{1–3} related to elliptic curves. The error terms in such problems hold the key to counting solutions of algebraic equations that have a wide range of applications, including cryptography and analyses of the speed of computer algorithms.

The proof came by combining some wonderful pieces of mathematics, and the key to it is all is so-called representation theory. This branch of mathematics, in its various guises, studies abstract groups by representing them as groups of linear transformations of vector spaces. By understanding the profound number-theoretic structure behind enough of the symmetric tensor powers of a certain representation of a certain group, one can compute the probability distribution of the corresponding scaled error terms, and so confirm the Sato–Tate conjecture.

The first article of the three, by Laurent Clozel, Michael Harris and Richard Taylor ('Automorphy for some l -adic lifts of automorphic mod l representations')¹, deals with the relationship between number theory and representation theory related to these symmetric

ANALYTICAL CHEMISTRY

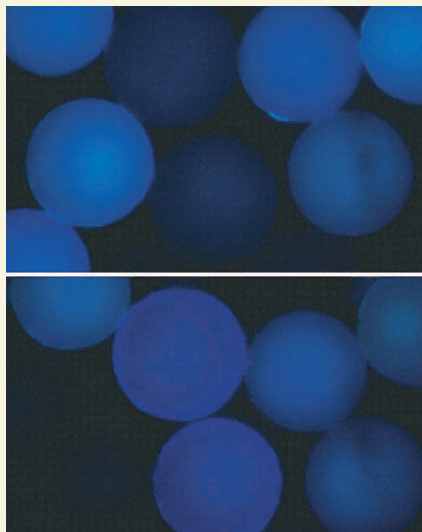
Playing molecular tag

With security tightened and concern about identity heightened, the computerized tagging of molecules with other molecules is an idea whose time has come.

A. Prasanna de Silva and colleagues present an appropriate strategy, which bears the grand title of molecular computational identification, or MCID (*Nature Mater.* doi:10.1038/nmat1733; 2006).

In essence, the technique involves using a series of single-input 'logic gates' — in this case, fluorescent dyes — that respond differently to various chemical stimuli, and whose combined response is unique to a particular molecule. That might be useful to those in the pharmaceutical industry, for example, who need to be able to keep close tabs on libraries of different, but structurally related, compounds.

The authors set about finding fluorescent molecules that reproduce the effects of five different logic gates when exposed to the H^+ ion, which is responsible for acidity in water. Two of these gates, PASS 1 and PASS 0, produce a positive or a negative



response regardless of the input and are trivial: they are, respectively, a fluorescent dye immobilized on a polymer bead, and no tag molecule at all.

For the YES gate, a molecule was chosen in which a process known as photoinduced electron transfer hinders fluorescence. When H^+ ions are present, they interact with electrons that would otherwise be transferred, so the molecule fluoresces. By using tag molecules with subtly different structures, the H^+ concentration at which this effect kicks in can be altered, and different colours of fluorescence can be created.

Similar principles apply to the authors' NOT gate, which

is activated when the concentration of H^+ is low. Pictured is an array of tag beads that has been immersed in hydrochloric acid (where the H^+ concentration is high; top image) or the alkali sodium hydroxide (with a lower level of H^+ ions; bottom image). The disappearance of two YES beads (bottom left) and the appearance of two NOT beads (middle) is clearly visible.

The technique can be easily extended to

multiple-input gates, as the authors demonstrate with a tag molecule that acts as an AND gate. It fluoresces only in the presence of both H^+ and the sodium ion Na^+ . Other distinguishable input stimuli are calcium, caesium and zinc ions, glucose molecules, and even heat and light.

The various available inputs, gates, switching thresholds and output colours could make MCID a sophisticated cataloguing tool, say the authors. Molecules and other nanoscale objects, like so much in today's world, may soon become familiar with computerized ID.

Richard Webb

tensor powers. Harris, Nicholas Shepherd-Barron and Taylor (in 'Ihara's lemma and potential automorphy')² then show that the necessary 'profound number-theoretic understanding' of the n th symmetric tensor power is, for even values of n , intimately connected to the algebraic geometry of a certain beautiful family of $(n-1)$ -dimensional vector spaces. This family, in the homogeneous variables $X_0, X_1, X_2, \dots, X_n$, is defined by the equation

$$X_0^{n+1} + X_1^{n+1} + X_2^{n+1} + \dots + X_n^{n+1} = tX_0X_1X_2\dots X_n,$$

where t is a parametrization variable. When $n=2$, this family — known as the Hessian — already played an important role in nineteenth-century geometry. The higher examples ($n>2$) are well known to theoretical physicists because of their relevance to mirror symmetry and conformal field theory, theories that are important in string theory and statistical mechanics. Finally, Taylor (in 'Automorphy

for some l -adic lifts of automorphic mod l representations II')³ overcame the last obstacle, providing a striking argument that links this 'intimate connection'² to the truth of the Sato–Tate conjecture in the class of problems mentioned above.

This is a magnificent achievement for at least two reasons. First, the method brings synthetic unity to deep results in quite distinct mathematical fields. This coming together is as startling as the theory of continental drift that connects the shape of disparate continents.

Second, the work answers a question of delicate nature. Number theorists have long held the opinion that the 'error terms', despite the pejorative name, have a mesmerizingly rich structure (they are the Fourier coefficients of fascinating mathematical objects known as cusp forms) and that the keys to some of the deepest issues in their subject lie hidden in that structure. ■



50 YEARS AGO

'Transplantation immunity' may be defined as the state of acquired resistance which leads to the destruction of homografts, that is, grafts transplanted between two different members of the same species... It is widely believed, for want of clear evidence to the contrary, that only living cells can elicit transplant immunity... It will be shown here that the power of cells to elicit transplantation immunity depends neither upon their being alive nor upon their structural integrity in any anatomical sense; and that the antigens responsible for skin transplantation immunity are certainly nuclear substances, and probably deoxyribonucleoproteins. From *Nature* 8 September 1956.

100 YEARS AGO

Another discussion of great interest, entitled "The Physiological Value of Rest," was introduced by Dr. Theodore Dyke Acland and Dr. Bevan Lewis... [they dealt with] the necessity for limiting the prevalent idea that "recreation is a change of occupation." This dictum is useful and true so long as occupation does not amount to fatigue, but its utility ends at this point. When the system becomes fatigued, and this is especially true of the brain, the toxic bodies produced affect unused as well as used cells. It is futile to throw these cells, already prejudiced, into activity. Such action simply adds to the amount of poisonous or toxic bodies in circulation. This point was worked out with great clearness by Dr. Bevan Lewis... The practical outcome of his argument, as well as of Dr. Acland's, was that physical exercise was no substitute for sleep, but that active physical exertion added to severe mental strain demanded a double need of slumber. In illustration of this point Dr. Acland recounted how Mr C. B. Fry, at once a scholar and an athlete, frequently slept till midday or even late in the afternoon during his school vacations, and in doing so gratified nothing more than the healthy demand of his frame — physical and mental — for rest. From *Nature* 6 September 1906.

Barry Mazur is in the Department of Mathematics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA. e-mail: mazur@math.harvard.edu

1. Clozel, L., Harris, M. & Taylor, R. www.math.harvard.edu/~rtaylor (2006).
2. Harris, M., Shepherd-Barron, N. & Taylor, R. www.math.harvard.edu/~rtaylor (2006).
3. Taylor, R. www.math.harvard.edu/~rtaylor (2006).
4. www.bts.gov
5. Hasse, H. *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* **3**, 253–262 (1933).
6. Weil, A. *Proc. Natl Acad. Sci. USA* **27**, 345–347 (1941).
7. Deligne, P. *Publ. Math. IHES* **43** (Presses Univ. France, Paris, 1974).
8. Tate, J. in *Arithmetic Algebraic Geometry* 93–110 (Harper & Row, New York, 1963).
9. Jackson, A. *Not. Am. Math. Soc.* **50**, 569–570 (2003).
10. Stein, W. <http://sage.math.washington.edu/sato-tate>

ORGANIC CHEMISTRY

Catalysts break symmetry

Scott E. Denmark

The creation of asymmetric molecules from symmetrical precursors is a useful strategy for organic synthesis. A new catalyst can accomplish this task through a unique, symmetry-breaking reaction.

Chirality, the 'left-handed' or 'right-handed' property of objects that are mirror images of each other, is ubiquitous in the observable world — shoes, conch shells, wood screws and umbilical cords, for instance. This property extends to objects in the nanoscale dimension; nearly all molecules in nature (such as amino acids, sugars, alkaloids and terpenes) and legions of synthetic compounds are chiral. For chemists who wish to prepare purpose-built compounds with specific attributes — for example, medicinal, mechanical, physical — the ability to imbue molecules with the appropriate chirality is paramount. In this issue, Zhao *et al.*¹ (page 67) describe a method that allows chiral molecules to be generated from achiral precursors. The chiral products are of great use as general building-blocks for further organic synthesis.

Chirality is fundamental to the structure, properties and function of molecules. How such molecules interact with one another in recognition, binding and chemical reaction is

crucially dependent on their chirality.

The synthesis of one chiral form of a molecule — known as a single enantiomer — can be achieved by the application of several well-established principles². The simplest method is to physically separate equal mixtures of enantiomers, a process known as resolution. Another powerful method is asymmetric synthesis, in which chiral molecules are prepared from achiral precursors by the action of a chiral reagent or catalyst³. In fact, nature has evolved highly selective enzymes to construct the chiral molecules of life, and these have been adapted for laboratory applications with unnatural substrates⁴.

An interesting subclass of asymmetric synthesis involves reactions that select between enantiotopic groups in achiral molecules with reflection symmetry (Fig. 1). These groups are indistinguishable except in a chiral environment or when acted upon by a chiral agent. A reaction at one of the enantiotopic groups will produce a chiral molecule, whereas the

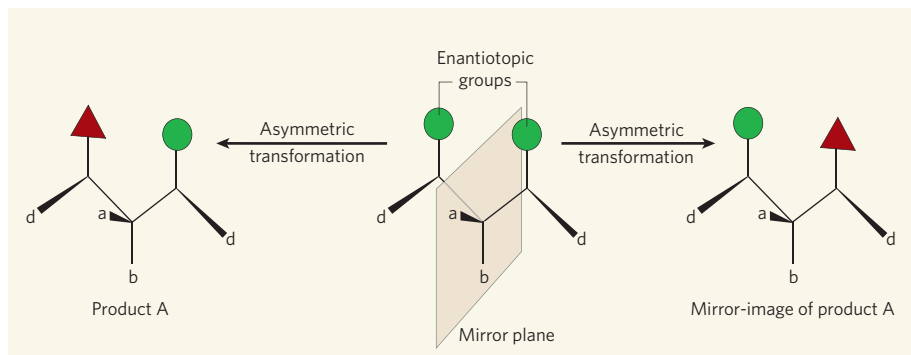


Figure 1 | Symmetry breaking. The central molecule contains a mirror plane of symmetry; **a**, **b** and **d** represent any chemical group. The spherical groups are enantiotopic — they differ by being on opposite sides of the plane of symmetry, but are otherwise identical. Enantiotopic groups can be distinguished by a chemical transformation with a chiral reagent, indicated by the conversion of a spherical group to a triangular group. The products of this transformation are not the same — they are mirror-images, and are said to be enantiomers of each other. The enantiomeric form of the product depends on which enantiotopic group is transformed. Zhao *et al.*¹ describe a catalyst that can distinguish between enantiotopic groups in a molecule.

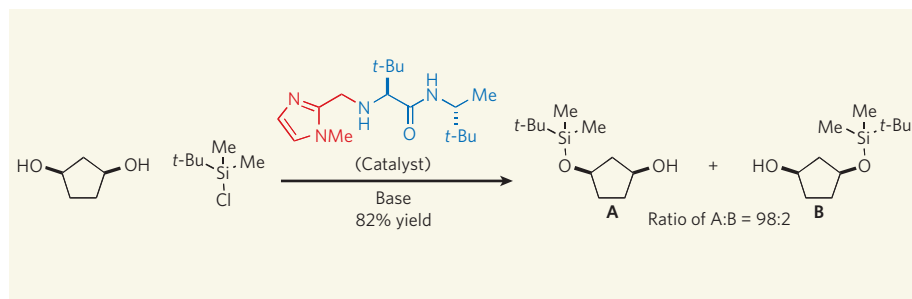


Figure 2 | Catalytic discrimination. A symmetrical molecule containing two hydroxyl (OH) groups reacts selectively at one of the groups with a silyl chloride under the influence of the catalyst reported by Zhao *et al.*¹, to yield predominantly the product A, rather than its mirror image, B. The catalyst contains two key subunits, the catalytic imidazole group (red) and the dipeptide molecular recognition element (blue). The silyl chloride contains two methyl groups (Me) and a bulky hydrocarbon group (*t*-Bu). Bold lines represent bonds to groups above the plane of the paper.

same reaction at its counterpart will produce the opposite enantiomeric product. Certain enzymes⁵ and a few synthetic organic catalysts^{6–8} can recognize and functionalize enantiotopic groups in molecules.

Zhao *et al.*¹ have now devised a non-biological catalyst that can effect this transformation with synthetically useful trialkylsilyl groups. Temporarily attaching trialkylsilyl groups to sensitive portions of organic molecules is a common strategy to prevent unwanted transformations of reactive chemical groups^{9,10}. The team has amalgamated the utility of the temporary protection strategy with the symmetry-breaking construction of value-added, chiral building-blocks in single enantiomer form.

The authors¹ surveyed a large number of small molecules (with a molecular weight of approximately 300 g mol⁻¹) as potential catalysts for this transformation. To select for one of the two hydroxyl (OH) groups in a simple test substrate, two key components were identified: a molecular recognition element (a dipeptide based on the amino acid *L*-*tert*-leucine) and a group to accelerate the reaction (an imidazole unit; Fig. 2). After optimization of the catalyst structure and reaction conditions, the authors settled on a specific catalyst and one of the most common trialkylsilyl precursors, *tert*-butyldimethylsilyl chloride (although others were shown to be suitable as well).

They then applied the optimized reaction to a myriad of substrates that contained hydroxyl groups in different structural settings. The chemical yields and enantiomeric selectivities of the reactions are very good — in some cases, excellent — although the reaction times are long (2–3 days). Relatively large amounts of the catalyst are required, but this is not a problem as the catalyst is simple to prepare from inexpensive starting materials. Clearly this is just the beginning of a development process and more active catalysts will be forthcoming.

The chiral catalyst¹ is remarkable because this transformation lacks any biological precedent. Nevertheless, the underlying concept of this catalytic process has an excellent chemical precedent known as Lewis base catalysis¹¹. The precursor of the trialkylsilyl protecting

group is activated by association with the imidazole group in such a way as to render it more prone to react with a given hydroxyl group. By embedding the imidazole in a specific dipeptide, the activated silyl–imidazolyl complex resides in a chiral environment. This enables the complex to differentiate between the two enantiotopic hydroxyl groups, reacting with one of them selectively and more rapidly than in the absence of the catalyst.

Although this work¹ concerns the preparation of mono-protected chiral diols — molecules that contain two hydroxyl groups — from symmetric starting materials, more general applications may be envisaged. For example, hydroxyl groups in similar chemical environments within asymmetric diols might also be selectively protected in this way.

The procedure might also be useful for polyols — molecules that contain more than two hydroxyl groups.

Given the role of protecting groups in the synthesis of complex molecules for pharmaceutical, agrochemical, materials science and academic enterprises^{9,10}, this procedure is likely to have a significant impact on the efficiency and cost of constructing single-enantiomer products. Most importantly, however, this report demonstrates the creative power of synthetic chemistry to build simple organic catalysts that mimic and ultimately surpass, biological catalysts — especially for non-biological transformations. ■

Scott E. Denmark is in the Department of Chemistry, University of Illinois, 600 S. Mathews Avenue, Urbana, Illinois 61801, USA.
e-mail: denmark@scs.uiuc.edu

1. Zhao, Y., Rodrigo, J., Hoveyda, A. H. & Snapper, M. L. *Nature* **443**, 67–70 (2006).
2. Mulzer, J. in *Stereoselective Synthesis: Methods of Organic Chemistry (Houben-Weyl)* 21st edn, Vol. 1 (eds Helmchen, G. *et al.*) 75–146 (Thieme, Stuttgart, 1996).
3. *Comprehensive Asymmetric Catalysis Vols I–III* (eds Jacobsen, E. N., Pfaltz, A. & Yamamoto, H.) (Springer, Heidelberg, 1999).
4. Thayer, A. M. *Chem. Eng. News* **84**(33), 15–31 (2006).
5. Ohno, M. & Otsuka, M. *Org. React.* **37**, 1–55 (1989).
6. Poss, C. & Schreiber, S. L. *Acc. Chem. Res.* **27**, 9–17 (1994).
7. Deng, L. *Chem. Rev.* **103**, 2965–2984 (2003).
8. Jarvo, E. R. & Miller, S. J. *Tetrahedron* **58**, 2481–2495 (2002).
9. Kocienski, P. J. *Protecting Groups* (Thieme, Stuttgart, 2005).
10. Wuts, P. & Green, T. *Protecting Groups in Organic Synthesis* 3rd edn (Wiley-Interscience, New York, 2005).
11. Denmark, S. E. & Stavenger, R. A. *Acc. Chem. Res.* **33**, 432–440 (2000).

MOLECULAR BIOLOGY

Sticky end in protein synthesis

Hervé Roy and Michael Ibba

It's not clear what general level of accuracy is required in translating the genetic code. But the protective role of proof-reading is evident from a case in which a small mistake has a catastrophic effect.

When protein production in a cell goes awry, abnormal deposits can form and contribute to various diseases known collectively as amyloidoses¹. The tendency of certain proteins to aggregate can be increased by mutations in the genes that encode them, as in Huntington's or Alzheimer's diseases. Or aggregation can be induced by an external factor such as a prion, for example in conditions such as mad cow disease or kuru.

On page 50 of this issue, Lee *et al.*² describe another pathway by which pathological cellular aggregates can form*. Certain mutant mice, called 'sticky' mice because of the effects on their fur, also suffer from neurodegeneration

*This article and the paper concerned² were published online on 13 August 2006.

caused by the death of a type of brain cell — the Purkinje cell — induced by protein aggregation. In investigating the origins of this neuron death, Lee *et al.* unexpectedly identified a mutation in the gene coding for alanyl-tRNA synthetase (AlaRS), a component of the gene-expression machinery by which aminoacyl-transfer RNAs translate messenger RNA into an amino-acid sequence on the ribosome. Closer inspection revealed that this mutation increases the frequency of errors during translation, leading to the gradual accumulation of inaccurately synthesized proteins that eventually form aggregates. As well as discovering another cause of cellular protein aggregation, Lee *et al.* demonstrate the importance of quality control by the amino-acyl-tRNA synthetases

(aaRSs), the family to which AlaRS belongs, in the intricate context of a multicellular organism.

The fidelity of information transfer is carefully controlled at each step during gene expression (Fig. 1). The most error-prone steps occur at the ribosome, where an aminoacyl-tRNA is matched with the corresponding codon (specifying an amino acid) on the messenger RNA, and an amino acid is paired with a tRNA by an aaRS³. With a few exceptions⁴, mistakes from the aminoacylation reaction cannot be corrected, which almost inevitably leads to the incorporation of the wrong amino acid during protein synthesis⁵. AlaRS is the aaRS that attaches the amino acid alanine (Ala) to its corresponding transfer RNA (tRNA^{Ala}) during protein synthesis.

The mutation Lee *et al.* identified does not impair the expression, structure or solubility of AlaRS, nor does it reduce the capacity for Ala-tRNA^{Ala} synthesis; instead, the mutated protein is less able to recognize and correct its own mistakes, leading to the accidental attachment of serine to tRNA^{Ala}. This is because the mutation leads to a sequence impairment in the AlaRS 'editing site', the domain of the protein that normally ensures that incorrectly activated amino acids are removed before they can be used for protein synthesis⁶. Editing by aaRSs is found in numerous systems, and serves to eliminate errors that inevitably arise when trying to discriminate between pairs of similar amino acids⁷.

In the case of AlaRS, serine is mistakenly activated and attached to tRNA at the active site with a frequency of about 1 in 500 compared with alanine. Then, rather than being released for protein synthesis, as occurs for alanine, the misactivated serine moves to the editing site where it is removed from the tRNA. It is this final step that is specifically impaired in the sticky mouse; Lee and colleagues' *in vitro* analyses² showed that, rather than hydrolysing Ser-tRNA^{Ala}, the mutated AlaRS produces more of this mischarged species, leading to errors in protein synthesis when alanine codons are mistakenly translated as serine. The results of these mistakes, the frequency of which is unknown, are protein misfolding, accumulation of protein aggregates and progressive neurodegeneration.

Editing by aaRSs has long been assumed to be a point of quality control in translation, although evidence for this role has been sparse until now. Studies in microorganisms showed that defects in editing impaired competitiveness but were not usually lethal^{8,9}, suggesting that cells can cope with a certain degree of error at this final step of gene expression. The work of Lee *et al.*² shows just how fine the line is between a tolerable degree of error and a catastrophic loss of accuracy in protein synthesis. The loss in editing activity caused by the mutated AlaRS has no discernible effect on the efficiency of protein synthesis in non-neuronal cells, but has a detrimental effect on Purkinje

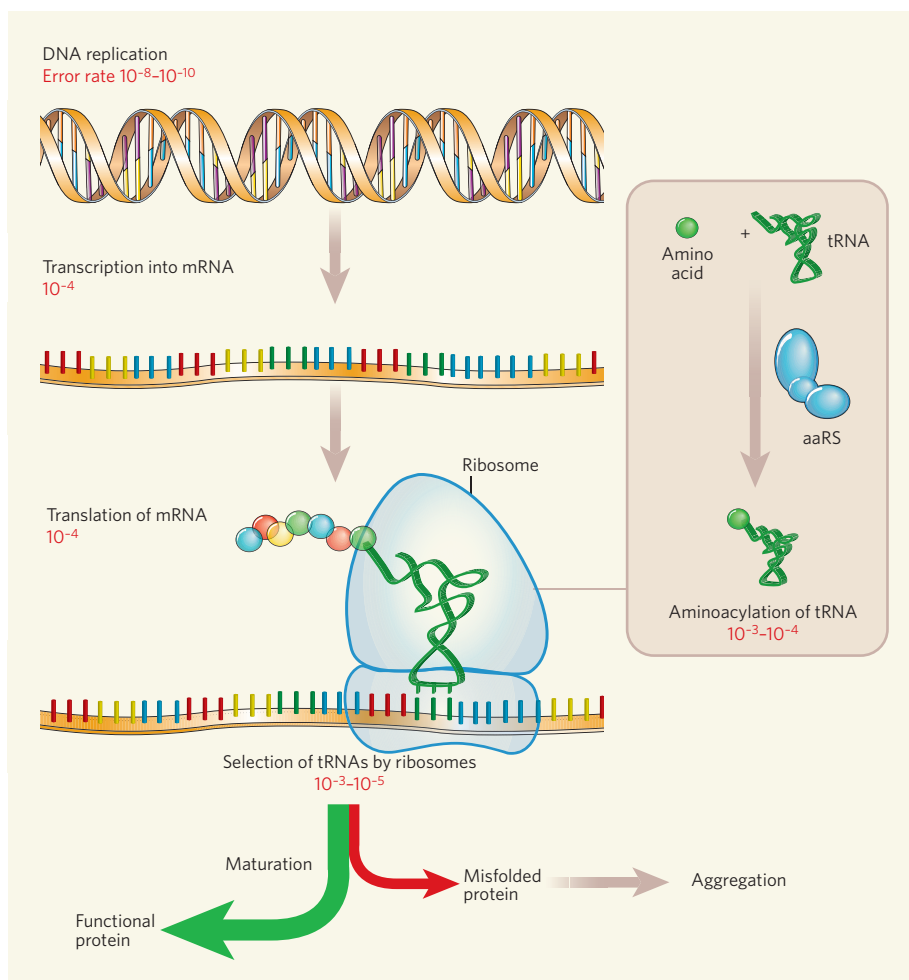


Figure 1 | Accuracy in gene expression. Transfer of information from the DNA sequence of a gene to the corresponding amino-acid sequence of a protein requires transcription of the sequence into messenger RNA, then translation of that RNA into an amino-acid sequence at ribosomes, through the agency of amino-acid-specific transfer RNAs. These steps are all prone to error (typical rates are shown in red). The inset shows the aminoacylation of tRNA, in which an amino acid is paired with a tRNA by the enzyme aminoacyl-tRNA synthetase (aaRS). It is a failure of quality control at this stage that Lee *et al.*² show causes neurodegeneration in 'sticky mice'. Once synthesized, a protein usually becomes functional. In some cases, however, misfolding occurs. Most misfolded proteins are degraded by the cell — but some form insoluble aggregates that, as in the sticky mouse, can lead to disease.

cells. This may be because, as these cells do not divide, there is no 'dilution' of the misfolded proteins by cell division. One question not yet addressed is whether the sticky-mouse characteristics arise from elevated serine misincorporation during synthesis of a particular subset of proteins in Purkinje cells, or from a general loss of fidelity in translation.

The discovery that 'upstream' defects in protein synthesis might lead to amyloidoses could open new routes for treating these devastating diseases. For example, just as Lee *et al.* show that in mutant mouse cells elevated serine levels decrease viability, one can naively speculate that a serine-restricted diet might limit erroneous protein synthesis to a level that neuronal cells can tolerate. Clearly, this is an oversimplification, and much remains to be learned both about the formation of cellular protein aggregates and how failures in quality control contribute to this and perhaps other diseases. However that turns out, it is clear that

the margin for error in translation is smaller than we thought.

Hervé Roy and Michael Ibba are in the Department of Microbiology and the Ohio State Biochemistry Program, Ohio State University, Columbus, Ohio 43210-1292, USA.
e-mail: ibba.1@osu.edu

- Merlini, G. & Bellotti, V. N. *Engl. J. Med.* **349**, 583–596 (2003).
- Lee, J. W. *et al.* *Nature* **443**, 50–55 (2006).
- Ibba, M. & Söll, D. *Science* **286**, 1893–1897 (1999).
- Dale, T. & Uhlenbeck, O. C. *Trends Biochem. Sci.* **30**, 659–665 (2005).
- Wang, L., Xie, J. & Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
- Beebe, K., Ribas De Pouplana, L. & Schimmel, P. *EMBO J.* **22**, 668–675 (2003).
- Jakubowski, H. & Goldman, E. *Microbiol. Rev.* **56**, 412–429 (1992).
- Nangle, L. A., De Crécy-Lagard, V., Döring, V. & Schimmel, P. *J. Biol. Chem.* **277**, 45729–45733 (2002).
- Korencic, D. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 10260–10265 (2004).

BRIEF COMMUNICATIONS

Nitrogen balance and Arctic throughflow

Waters moving east through the Arctic Ocean significantly contribute to nitrogen fixation in the Atlantic.

The Atlantic Ocean contains organisms that fix nitrogen, using up phosphate in the process and causing the waters to be enriched in nitrate relative to phosphate. The balance is redressed by the relatively phosphate-rich waters that flow into the Atlantic from the Arctic Ocean, through Fram Strait and the Canadian Archipelago. Here we show that the phosphate in this throughflow accounts for 16% or more of the nitrogen fixation in the North Atlantic. This unexpectedly high contribution highlights the importance of the Arctic throughflow in preventing large swings in nutrient budget and productivity in the ocean.

Nitrogen and phosphorus compounds regulate the productivity of the ocean¹ and therefore have an impact on the global carbon cycle. Both nutrients are supplied to the ocean by river run-off and by airborne dust, and are removed by sedimentation. Nitrogen can also be made biologically available by nitrogen-fixing marine organisms, or released back into the atmosphere as a result of denitrification, a process in which bacteria use nitrate to oxidize organic matter. Both processes take place in the Pacific and Atlantic oceans, with the Pacific providing a net sink of nitrogen and the Atlantic a net source², but how these are balanced is not well understood.

The Arctic Ocean receives water from the Pacific Ocean, which is already depleted in nitrate with respect to phosphate, through the Bering Strait (Fig. 1a, b). The broad shelves of the Bering and Chukchi seas are also denitrification sites^{3,4}, so water of Pacific origin is further reduced in nitrate relative to phosphate during its journey: we calculate that about $0.8 \mu\text{mol kg}^{-1}$ of phosphate remains after all the nitrate has been removed (Fig. 1a). This excess phosphate is transported into and through the Arctic Ocean, as shown by a shift in the nitrogen:phosphate ratio in water that is leaving through the Fram Strait and the Canadian Archipelago^{5,6}.

Using a Pacific inflow⁷ of 0.8 Sverdrup circulation units ($10^6 \text{ m}^3 \text{ s}^{-1}$), we estimate that 2×10^{10} mol excess phosphate is transported each year into the North Atlantic. Here, nitrogen-fixers use this excess phosphate and release nitrogen upon decomposition — initially in the form of ammonium — at a nitrogen:phosphorus ratio of about 45:1 (ref. 8). This enables organisms that are not nitrogen fixers to use phosphate with the newly fixed nitrogen at a ratio of 16:1. Under steady-state conditions, we

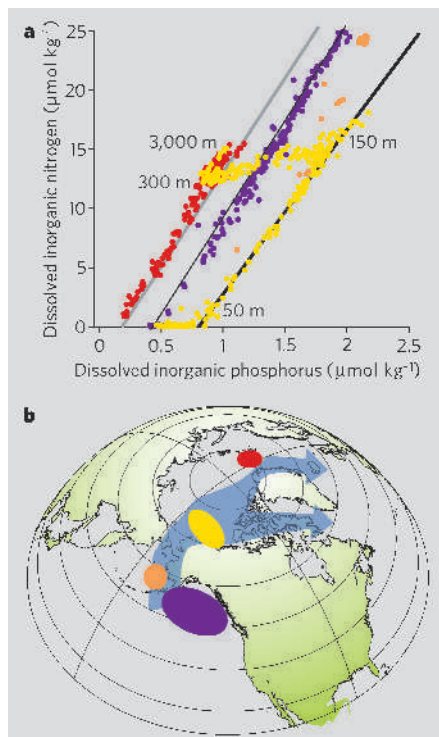


Figure 1 | Nitrogen balance between the Pacific and North Atlantic oceans. **a**, Relationship between dissolved inorganic nitrogen and dissolved inorganic phosphorus in waters in regions identified by colour on the map in **b**. Data are from the World Ocean Database 01 (red and purple), and from JAMSTEC (RV *Mirai* expeditions in 2000 and 2002; orange and yellow, respectively). The grey, thin black and thick black lines are regression lines for the uppermost 150 metres of the global ocean², northeastern Pacific waters and Canada Basin, respectively. Dissolved inorganic nitrogen comprises nitrates, nitrites and ammonium for the *Mirai* data set, and nitrates for the other data sets. Numbers indicate approximate depths in the Canada Basin. *Mirai* data were provided by T. Takizawa, K. Shimada, A. Murata and S. Nishino. **b**, Map showing the pathway followed by waters of Pacific origin into the North Atlantic^{5,6} (blue arrows). Colours show areas from which data plotted in **a** originate.

estimate that nitrogen-fixers would consume about 16/45 of the excess phosphate from the Arctic throughflow to fix $4.5 \times 10^{12} \text{ g N yr}^{-1}$: this represents 16% of the total nitrogen fixation in the North Atlantic².

If nitrogen-fixers are transported to depth before decomposition, however, uptake of excess phosphate by other organisms will be

depressed and will allow more nitrogen-fixers to use the excess phosphate. Indeed, a higher nitrogen:phosphorus ratio is observed in sinking particles, including nitrogen-fixers, in the western subtropical Pacific⁸, and we calculate (by using 23.3 as the mean nitrogen:phosphorus ratio) that an excess of 2×10^{10} mol phosphorus could account for 23% of the total nitrogen fixation in the North Atlantic.

This perspective suggests that in glacial times, when the Bering Strait was closed, excess phosphate (which was less than it is now owing to lower denitrification⁹) remained in the Pacific. The strait is now open and, with climate warming, the productivity of the Bering and Chukchi seas may rise¹⁰, thereby increasing denitrification¹¹ and the transport of excess phosphate to the North Atlantic would enable a compensating increase in nitrogen-fixation to make a larger contribution to balancing the global ocean nitrogen cycle.

The role of the Arctic throughflow as a mediator between regions of denitrification and of nitrogen-fixation has until now not been specifically acknowledged in the calculation of nutrient budgets or in models of the global ocean^{2,12}.

Michio Yamamoto-Kawai, Eddy Carmack, Fiona McLaughlin

Department of Fisheries and Oceans, Institute of Ocean Sciences, Sidney, British Columbia V8L 4B2, Canada
e-mail: kawaim@pac.dfo-mpo.gc.ca

1. Tyrrell, T. *Nature* **400**, 525–531 (1999).
2. Gruber, N. & Sarmiento, J. L. *Global Biogeochem. Cycles* **11**, 235–266 (1997).
3. Devol, A. H., Codispoti, L. A. & Christensen, J. P. *Cont. Shelf Res.* **17**, 1029–1050 (1997).
4. Tanaka, T. et al. *Cont. Shelf Res.* **24**, 1271–1283 (2004).
5. Taylor, J. R., Falkner, K. K., Schauer, U. & Meredith, M. *J. Geophys. Res.* **108**, 3374 (2003).
6. Jones, E. P. et al. *J. Geophys. Res.* **108**, 3116 (2003).
7. Woodgate, R. A. & Aagaard, K. *Geophys. Res. Lett.* **32**, L02602 (2005).
8. Karl, D. et al. *Biogeochemistry* **57/58**, 47–98 (2002).
9. Ganeshram, R. S. et al. *Paleoceanography* **15**, 361–376 (2000).
10. Loeng, H. et al. in *Arctic Climate Impact Assessment* (eds Symon, C. et al.) 453–538 (Cambridge Univ. Press, 2005).
11. Middelburg, J. J., Soetaert, K., Herman, P. M. & Heip, C. H. R. *Global Biogeochem. Cycles* **10**, 661–673 (1996).
12. Moore, J. K., Doney, S. C. & Lindsay, K. *Global Biogeochem. Cycles* **18**, GB4028 (2004).

Received 29 June; accepted 17 August 2006.
Competing financial interests: declared none.
doi:10.1038/443043a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

PLANETARY SCIENCE

Bedrock formation at Meridiani Planum

Arising from: T. M. McCollom & B. M. Hynek *Nature* 438, 1129–1131 (2005)

The Mars Exploration Rover Opportunity discovered sulphate-rich sedimentary rocks at Meridiani Planum on Mars, which are interpreted by McCollom and Hynek¹ as altered volcanic rocks. However, their conclusions are derived from an incorrect representation of our depositional model^{2,3}, which is upheld by more recent Rover data^{4–7}. We contend that all the available data still support an aeolian and aqueous sedimentary origin for Meridiani bedrock.

The McCollom and Hynek model¹ is based on their Fig. 1 (reproduced here as Fig. 1a), a ternary diagram of molar silicon and aluminium (Si + Al), magnesium, calcium and iron (Mg + Ca + Fe), and sulphur (S), which compares the molar chemical composition of Meridiani bedrocks with that of typical martian basalts. The apices of the shaded region are intended by McCollom and Hynek¹ to represent our model^{2,3} of the components of the outcrop rock when it was first deposited: sulphates, haematite and the siliciclastic fraction. Because the data lie outside this region, they claim that our model is implausible. We would point out, however, that the nature of the siliciclastic fraction of the rock — that is, the portion that consists of neither sulphates nor haematite — is misrepresented in the shaded region in Fig. 1a.

Although the sulphate and haematite apices are properly placed, the siliciclastic apex is not. We have consistently stated that the siliciclastic fraction is derived from chemical weathering of a basaltic precursor material^{2,8} and not that it consists of basaltic clasts, as asserted by McCollom and Hynek¹. We proposed that aluminosilicate minerals and a non-aluminous silicate, possibly free silica, make up much of the siliciclastic fraction⁸. We have also emphasized that the chemical alteration occurred before the material was incorporated into the outcrop⁸.

After chemical alteration, the distribution of cations is likely to be different from those in basaltic clasts: the altered siliciclastic residues would be shifted towards the (Si + Al) corner of Fig. 1a, especially under acidic conditions when iron is particularly mobile^{9,10}. In principle, our model is consistent with the siliciclastic material plotting all the way to the (Si + Al) corner (Fig. 1b). But the key point is that our model requires the alteration to be sufficient for the data points to lie within the shaded region.

McCollom and Hynek point out that the data fall on a mixing line between basalt and sulphur, and suggest that the rocks formed when sulphur was added to basalt in a volcanic environment. But the data available to them at the time were from a single small outcrop examined at the beginning of the mission^{1,2}; more recent data⁴ show a very different trend,

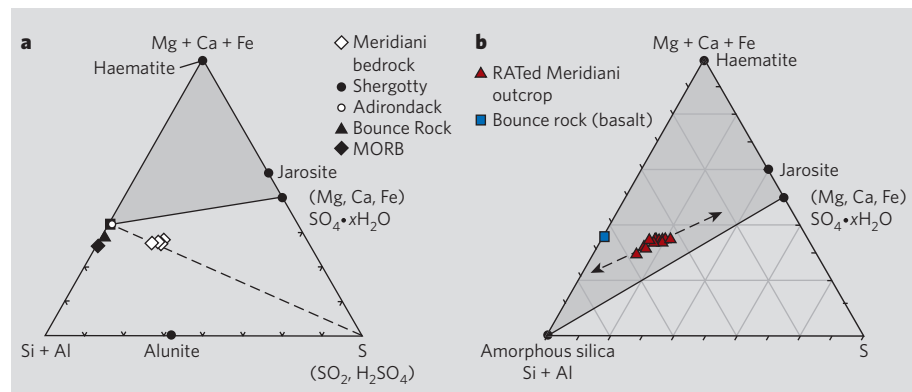


Figure 1 | Ternary diagrams of molar (Si + Al), (Mg + Ca + Fe) and sulphur. **a**, Ternary diagram taken from McCollom and Hynek¹. The shaded region is intended to represent our depositional model, but has one apex plotted for basalt, rather than for the altered basalt we advocated^{2–4,7–10}. **b**, Corrected diagram, in which the data points lie within the shaded region. In **a**, data points are shown only for a subset of our data that lie on a basalt–sulphur mixing line (dashed line). When all published data are included (**b**), a trend becomes evident that is inconsistent with that mixing line, but consistent with the mixing of altered basalt and sulphates plus haematite^{4,7–9}.

which can be explained instead by a mixture of altered basalt and another component consisting of sulphate salts and haematite (Fig. 1b).

Outcrop composition varies systematically with stratigraphic position: magnesium and sulphur co-vary and decrease with depth, whereas aluminium and silicon co-vary and increase with depth⁴. This variation is consistent with post-depositional interaction of the rocks with liquid water^{4–6,10}. For a volcanic model, however, it would require both an implausible vertical compositional gradient within the primary volcanic deposits and sulphur enrichment that correlates strongly with primary basalt composition, contrary to expectations for the interaction of sulphur-laden vapours with basalt.

Geological observations^{1,2,7} are also incompatible with a volcanic origin for Meridiani bedrock. The base surge deposits in Fig. 2b of ref. 1 are coarse-grained and poorly sorted, as are most base surge deposits. The rocks at Meridiani are uniformly fine-to-medium grained and are well sorted over the full area explored by the Rover^{2,7}. The cross-stratification shown by McCollom and Hynek¹ is an order of magnitude larger than that found at Eagle crater, and does not show the festoon geometry (nested, truncated, concave-upward geometry) observed at Meridiani⁷. On Earth, centimetre-scale festoon cross-lamination is known to develop only in subaqueous flows⁷.

Also, the broader stratigraphic context is neglected. Rocks at the Opportunity site form a 7-metre-thick set, subdivided into lower, middle and upper units, representing, respectively, aeolian dune (with internally complex cross-bed sets that are more than 2 m

thick), aeolian sand sheet, and mixed aeolian sand sheet and interdune facies associations⁷. They form a ‘wetting-upward’ succession that records a progressive increase in the influence of groundwater and, ultimately, surface water in controlling depositional processes⁷. The aeolian mixing has naturally led to chemistry that is laterally homogeneous over scales large and small. We know of no setting on Earth where interaction of sulphur-laden vapours with basaltic materials has produced a large volume of well mixed sulphate-rich sand grains. Moreover, a base surge origin requires a volcanic source topographically higher than the deposits themselves, yet there is no evidence for volcanoes anywhere on the Meridiani plain¹¹.

S. W. Squyres¹, O. Aharonson², R. E. Arvidson³, J. F. Bell III¹, P. R. Christensen⁴, B. C. Clark⁵, J. A. Crisp⁶, W. Farrand⁷, T. Glotch², M. P. Golombek⁶, J. Grant⁸, J. Grotzinger², K. E. Herkenhoff⁹, J. R. Johnson⁹, B. L. Jolliff², A. H. Knoll¹⁰, S. M. McLennan¹¹, H. Y. McSweeney¹², J. M. Moore¹³, J. W. Rice Jr⁴, N. Tosca¹¹

¹Department of Astronomy, Cornell University, Ithaca, New York 14853, USA

e-mail: squyres@astro.cornell.edu

²Division of Geological and Planetary Science, California Institute of Technology, Pasadena, California 91125, USA

³Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA

⁴Department of Geological Sciences, Arizona State University, Tempe, Arizona 85287, USA

⁵Lockheed Martin Corporation, Littleton, Colorado 80127, USA

⁶Jet Propulsion Laboratory, California Institute of

Technology, Pasadena, California 91109, USA

⁷Space Science Institute, Boulder, Colorado 80301, USA

⁸Center for Earth and Planetary Studies, Smithsonian Institution, Washington DC 20560, USA

⁹US Geological Survey, Flagstaff, Arizona 86001, USA

¹⁰Botanical Museum, Harvard University, Cambridge, Massachusetts 02138, USA

¹¹Department of Geosciences, State University of New York, Stony Brook, New York 11794, USA

¹²Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, Tennessee 37996, USA

¹³NASA Ames Research Center, Moffett Field, California 94035, USA

- McCollom, T. M. & Hynek, B. M. *Nature* **438**, 1129–1131 (2005).
- Squyres, S. W. *et al. Science* **306**, 1709–1714 (2004).
- Squyres, S. W. *et al. Science* **306**, 1698–1703 (2004).
- Clark, B. C. *et al. Earth Planet. Sci. Lett.* **240**, 73–94 (2005).
- McLennan, S. M. *et al. Earth Planet. Sci. Lett.* **240**, 95–121 (2005).
- Tosca, N. J. *et al. Earth Planet. Sci. Lett.* **240**, 122–148 (2005).
- Grotzinger, J. P. *et al. Earth Planet. Sci. Lett.* **240**, 11–72 (2005).
- Squyres, S. W. & Knoll, A. H. *Earth Planet. Sci. Lett.* **240**, 1–10 (2005).
- Nesbitt, H. W. & Wilson, R. E. *Am. J. Sci.* **292**, 740–777 (1992).
- Tosca, N. J., McLennan, S. M., Lindsley, D. H. & Schoonen, M. A. A. *J. Geophys. Res.* **109**, E05003, doi:10.1029/2003JE002218 (2004).
- Edgett, K. S. *Mars* **1**, 5–58 (2005).

doi:10.1038/nature05212

PLANETARY SCIENCE

McCollom & Hynek reply

Replying to: S. W. Squyres *et al. Nature* **443**, doi:10.1038/nature05212 (2006)

Squyres *et al.*¹ contend that our proposed volcanic origin for Meridiani Planum² is inconsistent with more recently obtained data³. But although the new data reveal some variation in chemical composition, this variation is small (Fig. 1a) and mainly due to modest variations in magnesium and sulphur, with concentrations of the other elements remaining essentially constant³. In a volcanic model, this variation can be readily explained by mobilization of highly soluble magnesium sulphate salts during the later stages of alteration and diagenesis (Fig. 1a), as in the sedimentary/evaporite model in which sediments that were initially deposited with uniform composition are subsequently modified^{3–5}. Although morphological features in the bedrock may be consistent with aeolian and fluvial origins⁶, this interpretation is not unique, particularly as features with similar grain size, sorting and morphology are seen in base surge deposits^{2,7–9}. Neither chemical nor morphological data therefore preclude a volcanic origin.

Squyres *et al.* claim¹ that their model is misrepresented in our Fig. 1 (ref. 2). However, the apices of the shaded triangle represent not the present composition of the mineral components, but the bulk composition of the potential primary chemical inputs in the sedimentary/evaporite model: a siliciclastic component, sulphate salts precipitated from evaporating groundwater, and iron that may have been mobilized to form haematite. In our Fig. 1 (ref. 2), the composition of the siliciclastic component was represented as basalt, consistent with their descriptions^{3–5,10}. Although weathering of silicate minerals is discussed^{4,5,10}, the current mineralogical composition of silicates in the bedrock places no definitive constraints on the chemical composition of the original siliciclastic component; this is because it is

inherently unclear whether the current minerals represent primary inputs or secondary alteration products. Consequently, the inferred presence of phyllosilicates and silica^{4,5,10} cannot be used reliably to constrain the bulk chemical composition of the original siliciclastic input.

Squyres *et al.*¹ suggest that we should have placed one apex of the shaded triangle at the Si + Al end-member (their Fig. 1b), but this is valid only if one of the primary chemical inputs had a bulk composition consisting of just Si + Al. However, the Si:Al ratio is constant throughout Meridiani bedrock³, and both the abundance of SiO₂ (48–53% by weight on a sulphur-free basis) and the Si:Al ratio (4.7–5.3) of the bedrocks are nearly identical to martian basalts¹¹ (48–51% by weight SiO₂; Si:Al, 3.5–7.2). There is thus no evidence for significant mobilization of Si or Al into the rocks, and Fig. 1b of Squyres *et al.*¹ does not accurately portray primary chemical inputs.

We agree that the current chemical composition of Meridiani bedrock can be accounted for by combining evaporitic and siliciclastic components¹². However, this siliciclastic component would have to have been substantially depleted in divalent cations and enriched in Si+Al relative to basalt before it ever interacted with evaporating fluids, not just before it was incorporated into the current outcrop (Fig. 1b). This critical requirement is not discussed in descriptions of the sedimentary/evaporite scenario^{3–5,10}, nor has a plausible source for such a large amount of material with this composition been proposed. Furthermore, it would be necessary for evaporating groundwater to add divalent cations to the cation-depleted siliciclastic material in the right proportions to result in their basalt-like chemical compositions¹², which seems improbable. The sedimentary/evaporite scenario has significant

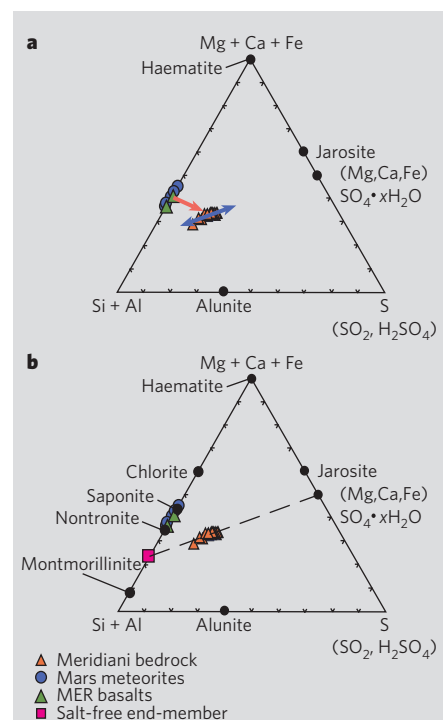


Figure 1 | Ternary diagrams showing relative molar abundances of major elements for Meridiani bedrock. Diagrams include data discussed by Squyres *et al.*¹ and typical martian basalts². **a**, In a volcanic scenario², bedrock compositions are attributable to reaction of basaltic ash with sulphuric acid from volcanic vapours. Minor scattering among compositions can be accounted for by mobilization of magnesium sulphate salts in the later stages of alteration (arrows). **b**, In a sedimentary/evaporite scenario, extrapolation from bedrock compositions to remove sulphate salts would require the original siliciclastic component to be substantially depleted in divalent cations and enriched in Si + Al relative to martian basalt¹².

shortcomings and alternative models^{2,7} need to be considered.

Thomas M. McCollom*†, Brian M. Hynek†

*Center for Astrobiology and †Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA
e-mail: mcollom@lasp.colorado.edu

- Squyres, S. W. *et al. Nature* **443**, doi:10.1038/nature05212 (2006).
- McCollom, T. M. & Hynek, B. M. *Nature* **438**, 1129–1131 (2005).
- Clark, B. C. *et al. Earth Planet. Sci. Lett.* **240**, 73–94 (2005).
- Squyres, S. W. & Knoll, A. H. *Earth Planet. Sci. Lett.* **240**, 1–10 (2005).
- McLennan, S. M. *et al. Earth Planet. Sci. Lett.* **240**, 95–121 (2005).
- Grotzinger, J. P. *et al. Earth Planet. Sci. Lett.* **240**, 11–72 (2005).
- Knauth, L. P., Burt, D. M. & Wohletz, K. H. *Nature* **438**, 1123–1128 (2005).
- Knauth, L. P., Burt, D. M. & Wohletz, K. H. *Lunar Planet. Sci. Conf. XXXVII abstr.* 1869 (2006).
- Burt, D. M., Wohletz, K. H. & Sheridan, M. F. *Lunar Planet. Sci. Conf. XXXVII abstr.* 2295 (2006).
- Squyres, S. W. *et al. Science* **306**, 1709–1714 (2004).
- Lodders, K. *Meteor. Planet. Sci.* **33**, 183–190 (1998).
- McCollom, T. M. & Hynek, B. M. *Lunar Planet. Sci. Conf. XXXVII abstr.* 2023 (2006).

doi:10.1038/nature05213

The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design

Rupert J. Russell^{1†}, Lesley F. Haire¹, David J. Stevens¹, Patrick J. Collins¹, Yi Pu Lin¹, G. Michael Blackburn², Alan J. Hay¹, Steven J. Gamblin¹ & John J. Skehel¹

The worldwide spread of H5N1 avian influenza has raised concerns that this virus might acquire the ability to pass readily among humans and cause a pandemic. Two anti-influenza drugs currently being used to treat infected patients are oseltamivir (Tamiflu) and zanamivir (Relenza), both of which target the neuraminidase enzyme of the virus. Reports of the emergence of drug resistance make the development of new anti-influenza molecules a priority. Neuraminidases from influenza type A viruses form two genetically distinct groups: group-1 contains the N1 neuraminidase of the H5N1 avian virus and group-2 contains the N2 and N9 enzymes used for the structure-based design of current drugs. Here we show by X-ray crystallography that these two groups are structurally distinct. Group-1 neuraminidases contain a cavity adjacent to their active sites that closes on ligand binding. Our analysis suggests that it may be possible to exploit the size and location of the group-1 cavity to develop new anti-influenza drugs.

Influenza virus membranes contain two glycoproteins: haemagglutinin and neuraminidase. Haemagglutinin mediates cell-surface sialic acid receptor binding to initiate virus infection. After virus replication, neuraminidase removes sialic acid from virus and cellular glycoproteins to facilitate virus release and the spread of infection to new cells¹. The distinct antigenic properties of different haemagglutinin and neuraminidase molecules are used to classify influenza type A viruses into subtypes: 16 for haemagglutinin (H1–H16) and 9 for neuraminidase (N1–N9)². Numerous combinations of haemagglutinin and neuraminidase subtypes are found in avian species; in humans, the three pandemics of the twentieth century were caused by viruses containing H1N1 in 1918, H2N2 in 1957 and H3N2 in 1968. The avian influenza virus that currently threatens a new pandemic is H5N1 (refs 3, 4). The N1 and N2 neuraminidases of viruses currently circulating in humans belong to two phylogenetically distinct groups⁵. Group-1 contains N1, N4, N5 and N8 subtypes whereas group-2 contains N2, N3, N6, N7 and N9 (Fig. 1a, b).

Neuraminidase has been targeted in structure-based enzyme inhibitor design programmes that have resulted in the production of two drugs, zanamivir (Relenza)⁶ and oseltamivir (Tamiflu)⁷, that to some extent mimic the transition state of the normal reaction. The success of these developments has been attributed, in part, to proposals that the catalytic sites of the enzymes are an invariant feature that might be exploited for subtype-independent therapy and to observations that they are comparatively rigid, with only minor conformational changes in the sites on inhibitor binding. Thus, the active sites of all influenza neuraminidases contain three arginine residues, Arg 118, Arg 292 and Arg 371, that bind the carboxylate of the substrate sialic acid; Arg 152 that interacts with the acetamido substituent of the substrate; and Glu 276 that forms hydrogen bonds with the 8- and 9-hydroxyl groups of the substrate. The X-ray crystallographic structural information that supports these conclusions is only available for the group-2 neuraminidases N2 and

N9 (refs 8, 9), but the idea that the active sites of the group-1 enzymes would be similar was supported by observations that the structures of more distantly related influenza type B neuraminidases¹⁰ are similar to those of the group-2 enzymes (Supplementary Fig. 1). Nevertheless, different drug-resistant neuraminidase mutant viruses have arisen after Tamiflu treatment of humans infected with viruses containing different neuraminidase subtypes^{11–14}. There are also indications that inhibitor structure/activity relationships do not apply across subtypes¹⁵. Given these anomalies, and the current concerns about the spread of avian H5N1 viruses¹⁶, we have determined the crystal structures of three group-1 neuraminidases from the N1, N4 and N8 subtypes and of their complexes with the inhibitors oseltamivir, zanamivir, DANA and peramivir, to compare their active sites with those of the group-2 enzymes against which the current drugs were designed^{8,9}.

The crystal structures of N1, N4 and N8 were solved by molecular replacement (see Methods and Supplementary Table 1).

Active site comparison

Superposition of the structures of N1, N4 and N8 group-1 neuraminidases reveals that their active sites are virtually identical (Fig. 1c, d). However, there are substantial conformational differences between group-1 and group-2 neuraminidases^{8,9,17} centred on the '150-loop' (residues 147–152) and the '150-cavity' adjacent to the active site (Fig. 1e). The conformation of the 150-loop is such that the C α position of group-1-specific Val 149 is about 7 Å distant from the equivalent isoleucine residue in group-2. Moreover, the hydrophobic side chain at position 149 is pointed away from the active site in group-1 but towards it in group-2. At the point of closest approach of the 150-loop to the active site, there is a difference of 1.5 Å in the side-chain position of the conserved aspartic acid residue at position 151 between group-1 and group-2 neuraminidases. A second nearby acidic residue, Glu 119, is also conserved and adopts a different conformation between

¹MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. ²Department of Chemistry, University of Sheffield, Sheffield S3 7HF, UK.

[†]Present address: Centre for Biomolecular Sciences, University of St Andrews, St Andrews KY16 9ST, UK.

the two groups. In group-2 structures this residue forms a hydrogen bond with Arg 156 but in group-1 it adopts a conformation such that its carboxylate points in approximately the opposite direction (Fig. 1e). However, comparison of the amino acid residues in the 150-loops offers no obvious explanation for the strong conservation of loop structure within, but not between, group-1 and group-2.

A major consequence of these differences in structure is that there is a large cavity adjacent to the active site in group-1 but not in group-2 neuraminidases (Fig. 2). This cavity is accessible from the active site because of the differences in position of Asp 151 and Glu 119 described above. The combined effect of the difference in position of these two acidic residues is to increase the width of the active site cavity by about 5 Å. The conserved Arg 156, the side chain of which is located approximately mid-way between the two acidic residues, adopts approximately the same position in the group-1 and group-2 structures and defines the entrance from the active site cavity into the 150-cavity. The extent of the 150-cavity is then determined by the difference in conformation of the 150-loop and by the position of Gln 136 (Fig. 2). In group-2 proteins this residue forms a hydrogen bond with the main-chain carbonyl of residue 150 of the loop. In group-1 structures, presumably as a consequence of the different loop structure, Gln 136, unable to make this hydrogen bond, adopts a conformation that results in its side chain sitting about 3.5 Å lower at the base of the cavity. The 150-cavity is therefore about 10 Å long and 5 Å wide and deep (Fig. 2).

The conformation of the active site residue Glu 276 has been of particular interest because it undergoes a marked rearrangement on

drug binding to neuraminidases from influenza B and group-2. For example, in unliganded group-2 (N9) the carboxylate of Glu 276 faces into the active site, but on oseltamivir binding it adopts a conformation pointing away from the active site so that the carboxylate now makes a bidentate interaction with the guanidinium group of Arg 224 (ref. 7). In the unliganded N1 (group-1) structure the conformation of Glu 276 (Fig. 1e) is very similar to that seen in unliganded N9 and, as we shall describe later, it undergoes a very similar rearrangement to that in N9 on oseltamivir binding.

Group-1 neuraminidase binding to inhibitors

We have determined the crystal structures of known anti-neuraminidase inhibitors in complex with N1, N4 and N8 (see Supplementary Table 1). Notably, we find that group-1 neuraminidases can bind oseltamivir in either the 'open' or 'closed' conformation of the 150-loop, depending on the soaking conditions. Thus, the structure of N8 neuraminidase in complex with oseltamivir, resulting from a 30-min soak of inhibitor into preformed crystals, reveals that no large-scale conformational changes have occurred (Fig. 3a) and that the 150-loop retains the same conformation as in the unliganded structure. Presumably as a consequence of the conformation of the 150-loop the acidic residues Asp 151 and Glu 119 are located further from the nitrogen attached to C4 of the inhibitor than they are in the complex with N9. Other interactions between oseltamivir and the N8 neuraminidase are similar to those observed in N9, with the further exception that Tyr 347 makes a hydrogen bond interaction with the C1 carboxylate of oseltamivir (Fig. 2) in addition to the usual

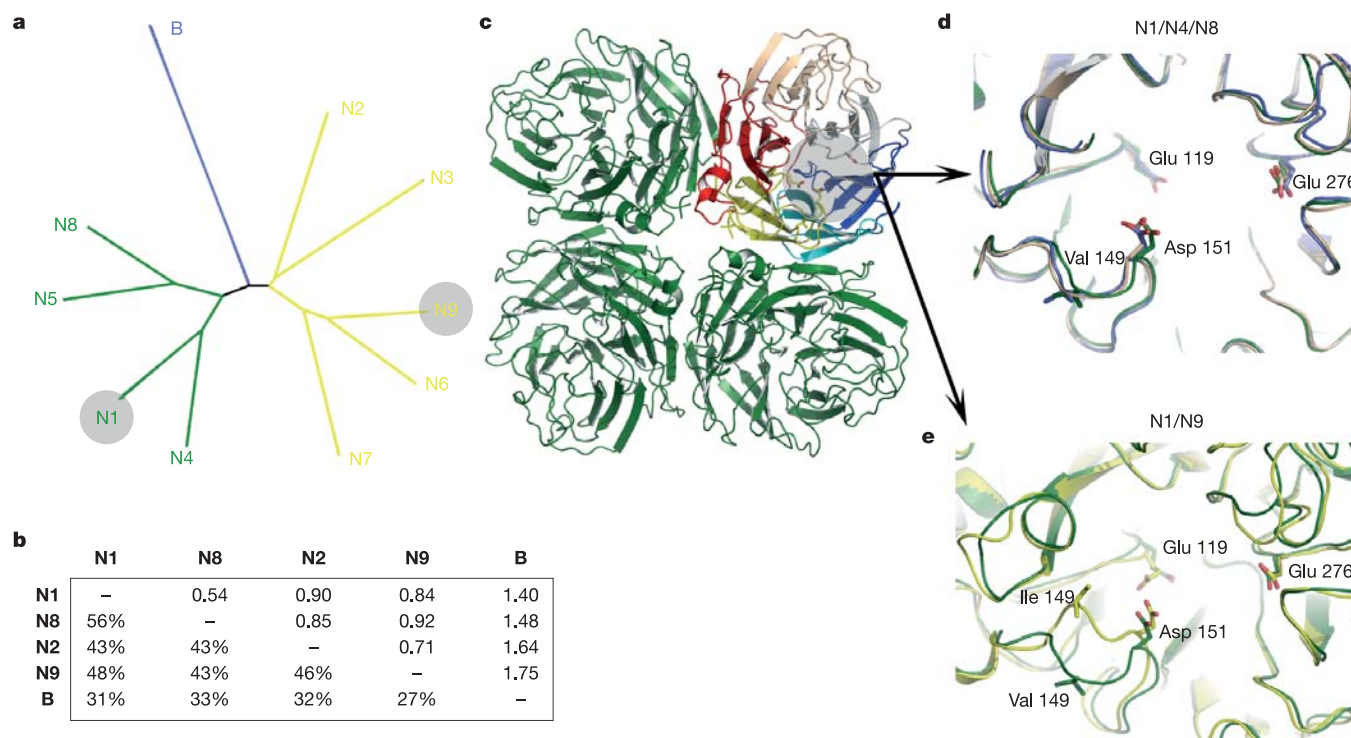


Figure 1 | Genetic and structural relationships between neuraminidases from different influenza viruses. **a**, Phylogenetic tree containing representative neuraminidases of influenza B and the nine subtypes of neuraminidase from influenza A. Influenza A neuraminidases fall into two distinct groups, which we have called group-1 and group-2 and coloured in green and yellow, respectively; the neuraminidase from influenza B is shown in blue. (Where appropriate, this colouring scheme is used in subsequent figures.) **b**, Table showing statistics for the similarities of sequence and structures between two members of group-1 (N1 and N8) and two from group-2 (N2 and N9) of influenza A with the neuraminidase from influenza B. The percentage sequence identities are shown on the left-hand side and root mean square deviations of C α positions between monomers is given on the right-hand side.

There is a strong correlation between the extent of sequence identity and the similarity of the crystal structures. **c**, Ribbons representation of the group-1 (N1) neuraminidase tetramer. One monomer is coloured to emphasize the molecules' canonical six-bladed β -propeller structure. The active site region at the centre of the six-bladed β -propeller structure is highlighted and then shown on a larger scale in **d** and **e**. **d**, Superposition of the active sites of three neuraminidases from group-1, showing how similar they are: N1, green; N4, gold; and N8, blue. **e**, Superposition of the active site of N1 (green) and N9 (yellow) neuraminidases, demonstrating that N9 is markedly different to N1 in the 150-loop region. Conserved residues such as Glu 119, Asp 151 and Glu 276 and the hydrophobic residue at position 149 are shown in stick representation.

bidentate interaction of that carboxylate with Arg 371. In group-2 neuraminidases, residue 347 is a glutamine, rather than a tyrosine, which is unable to make such a hydrogen bond.

It seems likely that the binding of oseltamivir to N8, at least in the crystalline state, is a two-step process. First, inhibitor binds to the 'open' form of N8 and then a slow conformational change occurs that results in the 'closed' form of the enzyme that is able to make a tighter interaction with ligand. At this stage we have no information about how the slow conformational change that occurs in the crystalline state relates to the enzyme in solution, but our structural observations show that this type of inhibitor is capable of binding to the open conformation of group-1 neuraminidases.

Incubating N1 crystals in 20 μ M oseltamivir for 150 min also showed binding of inhibitor with the 150-loop in the open cavity conformation (Supplementary Fig. 2), but when N8 crystals were incubated in oseltamivir for 3 days (Fig. 3a, b), or N1 crystals were incubated in a higher concentration of inhibitor (Fig. 3b), the 150-loop changes its conformation so that it closely resembles the conformation observed in group-2 neuraminidases in the presence

and absence of inhibitors. There are two main consequences of this change in conformation. First, Glu 119 and Asp 151 are now both oriented towards the bound oseltamivir, and second, the size of the active site cavity in drug-bound group-1 is now much the same as it is for group-2 neuraminidases.

We have also determined the structures of three other neuraminidase inhibitors, DANA^{6,18,19}, zanamivir²⁰ and peramivir²¹, bound to group-1 neuraminidases (Supplementary Fig. 3). These structures show that the drug-bound complexes of group-1 are very similar to those seen for group-2 neuraminidases.

Overall, the observation of the open conformation for the 150-loop in the group-1 structures suggests that, for these enzymes, this conformation of the loop is intrinsically lower in energy than the closed conformation. Group-1 neuraminidases (N1 and N8) initially bind to oseltamivir in this open conformation but eventually adopt the closed conformation. It thus seems that oseltamivir binding to group-1 neuraminidases favours the higher energy or closed conformation of the 150-loop that it probably accesses via a relatively slow conformational change. It should therefore be possible to design new inhibitors for group-1 neuraminidases that are selective for the open 150-loop conformation and would thereby have the potential to bind more strongly than oseltamivir or zanamivir.

Examination of the structures suggests, for example, that it may be possible for a new substituent to be developed from the 4-amino group of oseltamivir into the 150-cavity and thereby enhance the binding of potential inhibitors. The prominent guanidinium side chain of conserved Arg 156, at the base of the 150-cavity (Fig. 2), is

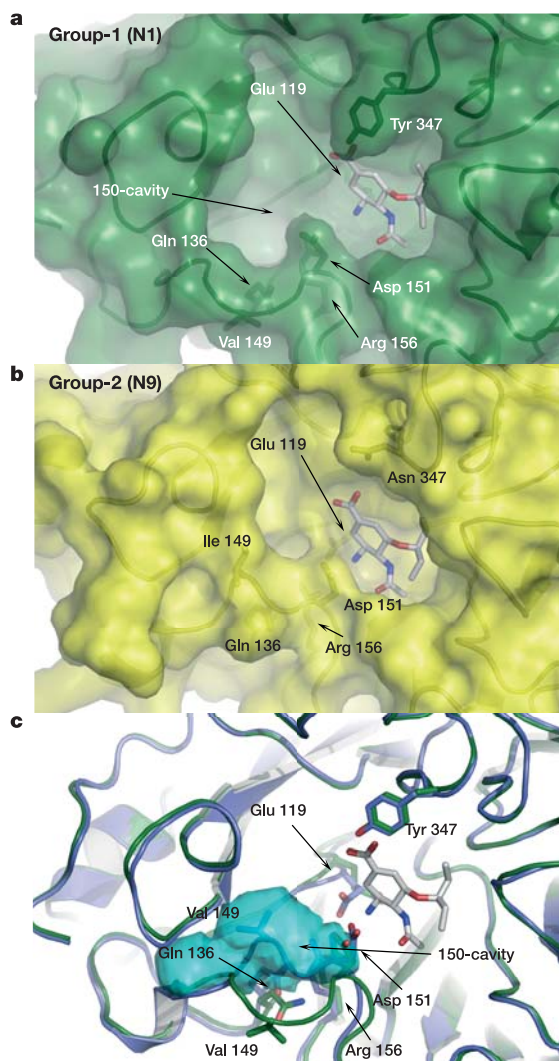


Figure 2 | Molecular surfaces of group-1 and group-2 neuraminidases with bound oseltamivir showing the 150-cavity in the group-1 structure that arises because of the distinct conformation of the 150-loop. a, b, N1 (a; green) and N9 (b; yellow) shown in surface representation with the protein main chain shown in 'worm' representation. c, Superposition of the active sites of apo-N1 (green) and N1 complexed with oseltamivir (blue). Part of the electron density map from a low-resolution (5.5 Å) difference Fourier calculated between apo-N1 and oseltamivir-bound N1 data sets is shown in blue to indicate the position of the 150-cavity.

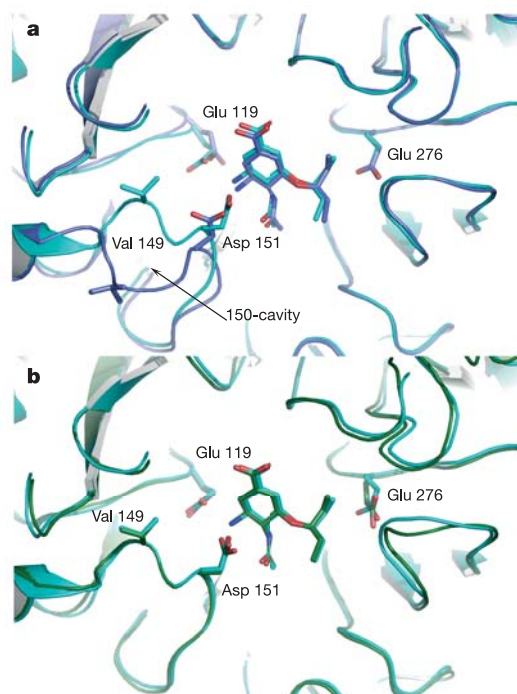


Figure 3 | Oseltamivir binding to the active sites of group-1 neuraminidases. a, Superposition of the active sites of N8 after a 30-min soak (dark blue) and a 3-day soak (cyan) with 20 μ M oseltamivir. There are small changes in the position of Glu 119 and the inhibitor when the 150-loop closes after the longer soaking time. b, Superposition of the active sites of N8 with bound oseltamivir after the 3-day soak with 20 μ M inhibitor (cyan) with N1 soaked for 30 min in 0.5 mM inhibitor (green). In this case, the structures of the two different subtypes of neuraminidase from group-1 are remarkably similar.

clearly a prospective partner for a salt-bridge or hydrogen bond with a new inhibitor.

Differential oseltamivir resistance of mutant neuraminidases

Three principal oseltamivir-resistant mutant neuraminidases have been characterized from influenza A viruses isolated after Tamiflu treatment of influenza-infected humans^{14,22}. One, derived from H1N1 and H5N1 infections, contained the amino acid substitution His274Tyr. The other two were from patients infected with H3N2 viruses and contained either Arg292Lys or, less frequently, Glu119Val substitutions. Although oriented differently in unliganded group-1 compared with group-2 neuraminidases, Glu119 of both groups appears to interact similarly with oseltamivir. Comparison of the structures of group-1 and group-2 neuraminidases reveals group-specific differences in the active sites that might explain how the mutations at positions 274 and 292 lead to inhibitor resistance.

The mutation His274Tyr leads to high resistance of group-1 neuraminidases against oseltamivir but has little effect on group-2 neuraminidases^{23,24}. Inspection of the structures of the group-1 neuraminidases in complex with oseltamivir, and comparison with equivalent group-2 complexes, suggests a reason for this group-specific difference and indicates how resistance may be mediated by

the effects of the mutant Tyr 274 on the orientation of Glu 276. The importance of the conformation of Glu 276 for oseltamivir binding by group-2 neuraminidases has been firmly established²⁵. There seems to be at least two factors contributing to the inability of group-1 neuraminidases to accommodate the His274Tyr substitution (Fig. 4). First, the 270-loop in group-1 neuraminidases approaching residue 273 makes a tighter turn than the equivalent loop in group-2. Second, in group-1 neuraminidases, but not in group-2, there is a conserved tyrosine residue at position 252 that makes hydrogen bonds to the main-chain carbonyl at position 273, to the peptide amide at 250 and to the histidine side chain at 274. Histidine 274 also forms a hydrogen bond through its other side-chain nitrogen with Glu 276. It appears that introduction of the bulkier tyrosine residue at position 274 in group-1 (N1) enzymes can only be accommodated by the new side chain moving towards, and partially displacing, Glu 276. By contrast, in group-2 enzymes, there is a smaller residue at position 252, leaving space for Tyr 274 to occupy without perturbing Glu 276. This interpretation of the group-specific effect of this mutation is consistent with observations from mutagenic studies that examined the effects of side-chain size at residue 274 on sensitivity to oseltamivir²³.

The mutation Arg292Lys is the commonest substitution in group-2 (N2) neuraminidases resistant to oseltamivir²². It has already been the subject of a detailed crystallographic analysis to show that in the group-2 neuraminidase N9, resistance results, in part, from the loss of a hydrogen bond from Arg292 to the carboxylate group of oseltamivir²⁵. The substituted Lys292 also interacts with Glu 276, impeding its movement to accommodate the hydrophobic substituent attached to C6 of oseltamivir. The structures of group-1 neuraminidases, and their complexes with oseltamivir, now reveal a likely reason for the smaller effect of the mutation on group-1 enzymes. As Fig. 4 shows, the conserved tyrosine residue at position 347 in the group-1 neuraminidase N1 makes an additional hydrogen bond to the carboxylate group of the inhibitor that cannot be made by the equivalent residues in group-2 neuraminidases. In this way it seems that the additional hydrogen bond interaction between Tyr 347 and the carboxylate of the inhibitor compensates for a weaker, water-mediated interaction between the carboxylate and the substituted lysine residue at position 292 (ref. 25).

Conclusion

As a proven anti-influenza drug target, neuraminidase continues to be attractive for the development of new virus inhibitors, not least because of the emergence of viruses resistant to the currently available drugs^{14,26–28}. The crystal structures of group-1 neuraminidases described here will add to this attraction. They show that the 150-loop, which forms one corner of the enzyme active site, is able to exist in at least two stable conformations. The fact that group-1 neuraminidases bind drugs like oseltamivir with similar affinity to group-2 enzymes²² suggests that the difference in energy between the two conformations is not very large. The notion of a degree of plasticity in the structure of the active site of neuraminidase, or at least of the group-1 enzymes, is unexpected, but considering their similarities in sequence it would not now be surprising if the 150-loop of group-2 neuraminidases was found to also possess a degree of flexibility. Evidently the closed conformation is energetically preferred in group-2 neuraminidases, both in the absence and presence of current inhibitors, but a higher energy open conformation may well be accessible to an inhibitor that could make an energetically advantageous interaction with it.

On the basis of our structural observations, new drugs against group-1 neuraminidases could be obtained by adding extra substituent moieties to existing inhibitor skeletons. Although it may be considered ideal to focus on compounds that work against all virus subtypes, an effective group-specific inhibitor could be of considerable value against the currently circulating human H1N1 viruses, the H3N8 viruses that repeatedly cause influenza in equines and are now

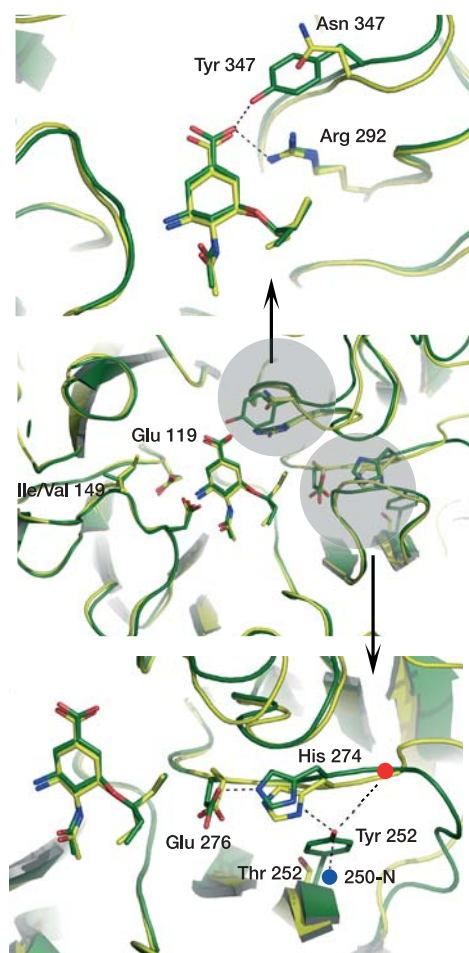


Figure 4 | Locations of the oseltamivir resistance mutations found in group-1 and group-2 neuraminidases. Middle panel: superposition of the active sites of group-1 (green) and group-2 (yellow) neuraminidases with bound oseltamivir; two regions of the active site are highlighted and detailed in the top and bottom panels. Top panel: residues involved in the group-2-specific resistance mutant Arg292Lys also showing that residue 347 is a tyrosine residue in the group-1 protein but an asparagine in group-2. Bottom panel: residues involved in the group-1-specific resistance mutant His274Tyr showing that the tyrosine at position 252 is involved in a network of hydrogen bonds in group-1.

causing widespread disease in canines²⁹, and the avian H5N1 viruses that currently threaten the human population. In any case, it may well be possible that new inhibitors designed to exploit additional interactions with the open form of the 150-loop of group-1 neuraminidases could select a similar conformation of this loop in group-2 neuraminidases.

METHODS

N1 neuraminidase (NA) was prepared from a WSN-NA (H5N1) recombinant virus containing seven genes from WSN and the neuraminidase gene from A/Vietnam/1203/04. N4 and N8 neuraminidases from A/mink/Sweden/E12665/84 (H10N4) and A/duck/Ukraine/1/63 (H3N8) were prepared from viruses grown in hens' eggs. Neuraminidase was released from the viruses by bromelain digestion, and further purified, as previously described³⁰. N1 crystals were grown by vapour diffusion in sitting drops dispensed by an Oryx-6 crystallization robot from Douglas Instruments. A total of 0.1 µl of protein solution, at 9 mg ml⁻¹ in 10 mM Tris-HCl, pH 8.0, was mixed with an equal volume of reservoir solution (0.1 M MES, pH 6.0, 0.2 M ammonium acetate, 20% w/v PEG 3350). N4 crystals were grown by vapour diffusion in hanging drops consisting of 2 µl of reservoir solution (0.1 M HEPES, pH 7.5, 5 mM cobalt chloride, 5 mM nickel chloride, 5 mM cadmium chloride, 5 mM magnesium chloride and 12% w/v PEG 3350) and 2 µl of concentrated protein solution (10 mg ml⁻¹ in 10 mM Tris-HCl, pH 8.0). N8 neuraminidase crystals were grown by vapour diffusion in hanging drops consisting of 2 µl of reservoir solution (0.1 M imidazole, pH 8.0 and 35% MPD) and 2 µl of concentrated protein solution (10 mg ml⁻¹ in 10 mM Tris-HCl, pH 8.0). Crystals of N4 and N8 were soaked for 30 min in inhibitor made up in crystallization buffer (augmented with 20% glycerol, in the case of N4, for cryo-protection) at concentrations of 20 µM for oseltamivir, zanamivir and peramivir and at 20 mM for DANA. Additionally, a crystal of N8 neuraminidase was soaked in 20 µM oseltamivir for 3 days. N1 was cryo-protected by the addition of 15% v/v ethylene glycol and the oseltamivir soaks were carried out using 20 µM or 0.5 mM drug made up in this cryo-protectant.

N4 and N8 data were collected at 100 K on an in-house Rigaku-MSU RU200 rotating anode coupled to a RaxisIIc detector. N1 data sets were recorded on a Raxis4 detector (100-µm scan) mounted on a Rigaku MicroMax 007 HF generator. Diffraction data were integrated using Denzo and scaled with Scalepack³¹. N1, N4 and N8 neuraminidase structures were solved by molecular replacement using variously AmoRe (N1) and Phaser³² (N4, N8) with N9 neuraminidase as the initial search model. Standard refinement was carried out with a combination of remlac³³ and CNS³³ together with manual model building with O³⁴. Figures were created with Pymol (<http://pymol.sourceforge.net/>).

Received 6 June; accepted 25 July 2006.

Published online 16 August 2006.

- Murphy, B. R. & Webster, R. G. In *Fields Virology* 3rd edn (eds Fields, D. B. N., Knipe, M. & Howley, P. M.) 1397–1445 (Lippincott-Raven, Philadelphia, 1996).
- World Health Organization. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. *Bull. World Health Organ.* **58**, 585–591 (1980).
- Bender, C. et al. Characterization of the surface proteins of influenza A (H5N1) viruses isolated from humans in 1997–1998. *Virology* **254**, 115–123 (1999).
- World Health Organization Global Influenza Program Surveillance Network. Evolution of H5N1 avian influenza viruses in Asia. *Emerg. Infect. Dis.* **11**, 1515–1521 (2005).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput. Appl. Biosci.* **10**, 19–29 (1994).
- von Itzstein, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418–423 (1993).
- Kim, C. U. et al. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **119**, 681–690 (1997).
- Varghese, J. N., Laver, W. G. & Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* **303**, 35–40 (1983).
- Baker, A. T., Varghese, J. N., Laver, W. G., Air, G. M. & Colman, P. M. Three-dimensional structure of neuraminidase of subtype N9 from an avian influenza virus. *Proteins* **2**, 111–117 (1987).
- Burmeister, W. P., Ruigrok, R. W. & Cusack, S. The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *EMBO J.* **11**, 49–56 (1992).
- Ives, J. et al. Anti-viral drug resistance: an oseltamivir treatment-selected influenza A/N2 virus with a R292K mutation in the neuraminidase gene has reduced infectivity *in vivo*. *J. Clin. Virol.* **18**, 251–269 (2000).
- Gubareva, L. V., Kaiser, L., Matrosovich, M. N., Soo-Hoo, Y. & Hayden, F. G. Selection of influenza virus mutants in experimentally infected volunteers treated with oseltamivir. *J. Infect. Dis.* **183**, 523–531 (2001).
- Carr, J. et al. Influenza virus carrying neuraminidase with reduced sensitivity to oseltamivir carboxylate has altered properties *in vitro* and is compromised for infectivity and replicative ability *in vivo*. *Antiviral Res.* **54**, 79–88 (2002).
- Ward, P., Small, I., Smith, J., Suter, P. & Dutkowski, R. Oseltamivir (Tamiflu) and its potential for use in the event of an influenza pandemic. *J. Antimicrob. Chemother.* **55** (suppl. 1), i5–i21 (2005).
- Brouillette, W. J. et al. Pyrrolidinobenzoic acid inhibitors of influenza virus neuraminidase: modifications of essential pyrrolidinone ring substituents. *Bioorg. Med. Chem.* **11**, 2739–2749 (2003).
- Webster, R. G., Peiris, M., Chen, H. & Guan, Y. H5N1 outbreaks and enzootic influenza. *Emerg. Infect. Dis.* **12**, 3–8 (2006).
- Colman, P. M., Varghese, J. N. & Laver, W. G. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **303**, 41–44 (1983).
- Bossart-Whitaker, P. et al. Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy-2,3-dehydro-N-acetyl neuraminic acid. *J. Mol. Biol.* **232**, 1069–1083 (1993).
- Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, S. & Ruigrok, R. W. Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure* **1**, 19–26 (1993).
- Meindl, P., Bodo, G., Palese, P., Schulman, J. & Tuppy, H. Inhibition of neuraminidase activity by derivatives of 2-deoxy-2,3-dehydro-N-acetylneuraminic acid. *Virology* **58**, 457–463 (1974).
- Babu, Y. S. et al. BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J. Med. Chem.* **43**, 3482–3486 (2000).
- Mishin, V. P., Hayden, F. G. & Gubareva, L. V. Susceptibilities of antiviral-resistant influenza viruses to novel neuraminidase inhibitors. *Antimicrob. Agents Chemother.* **49**, 4515–4520 (2005).
- Wang, M. Z., Tai, C. Y. & Mendel, D. B. Mechanism by which mutations at His274 alter sensitivity of influenza A virus N1 neuraminidase to oseltamivir carboxylate and zanamivir. *Antimicrob. Agents Chemother.* **46**, 3809–3816 (2002).
- Yen, H. L. et al. Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. *Antimicrob. Agents Chemother.* **49**, 4075–4084 (2005).
- Varghese, J. N. et al. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **6**, 735–746 (1998).
- Le, Q. M. et al. Avian flu: isolation of drug-resistant H5N1 virus. *Nature* **437**, 1108 (2005).
- Kiso, M. et al. Resistant influenza A viruses in children treated with oseltamivir: descriptive study. *Lancet* **364**, 759–765 (2004).
- de Jong, M. D. et al. Oseltamivir resistance during treatment of influenza A (H5N1) infection. *N. Engl. J. Med.* **353**, 2667–2672 (2005).
- Crawford, P. C. et al. Transmission of equine influenza virus to dogs. *Science* **310**, 482–485 (2005).
- Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl Acad. Sci. USA* **98**, 11181–11186 (2001).
- Otwinowski, Z. & Minor, W. In *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, 1993).
- CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brunger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the MRC (UK) and by an International Partnership Research Award in Veterinary Epidemiology from the Wellcome Trust. R.J.R. acknowledges the Wellcome Trust and the University of St Andrews for support and Biocryst for supply of peramivir. We thank P. Walker for assistance with data collection and preparation of the manuscript, and Rigaku (Europe) and CRUK (Lincoln's Inn) for access to data collection facilities.

Author Information Coordinates have been deposited with the Protein Data Bank and the relevant accession codes (2HTY, 2HU0, 2HU4, 2HT5, 2HTR, 2HT7, 2HT8, 2HTQ, 2HTU, 2HTV and 2HTW) are described in Supplementary Table 1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.J.S. (mbrenna@nimr.mrc.ac.uk).

ARTICLES

Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration

Jeong Woong Lee¹, Kirk Beebe², Leslie A. Nangle², Jaeseon Jang^{1†}, Chantal M. Longo-Guess¹, Susan A. Cook¹, Muriel T. Davisson¹, John P. Sundberg¹, Paul Schimmel² & Susan L. Ackerman^{1,3}

Misfolded proteins are associated with several pathological conditions including neurodegeneration. Although some of these abnormally folded proteins result from mutations in genes encoding disease-associated proteins (for example, repeat-expansion diseases), more general mechanisms that lead to misfolded proteins in neurons remain largely unknown. Here we demonstrate that low levels of mischarged transfer RNAs (tRNAs) can lead to an intracellular accumulation of misfolded proteins in neurons. These accumulations are accompanied by upregulation of cytoplasmic protein chaperones and by induction of the unfolded protein response. We report that the mouse sticky mutation, which causes cerebellar Purkinje cell loss and ataxia, is a missense mutation in the editing domain of the alanyl-tRNA synthetase gene that compromises the proofreading activity of this enzyme during aminoacylation of tRNAs. These findings demonstrate that disruption of translational fidelity in terminally differentiated neurons leads to the accumulation of misfolded proteins and cell death, and provide a novel mechanism underlying neurodegeneration.

The genetic code is established in the aminoacylation reactions of aminoacyl-tRNA synthetases, where each amino acid is linked to its cognate tRNA that bears the anticodon triplet of the code. The rate of misincorporation of amino acids into proteins is very low (estimated at one error in every 10^3 – 10^4 codons)^{1,2}, and this high accuracy results largely from the precision of aminoacylation reactions. In addition to tRNA recognition, aminoacyl-tRNA synthetases must discriminate between amino acids in the cellular pool. Generally, amino acids with side chains that are bulkier than those of the cognate amino acids are sterically excluded from the active sites of tRNA synthetases, but smaller amino acids can fit into the active site pocket and be misactivated and mischarged. These misactivated adenylates or mischarged tRNAs are normally cleared by the editing function of aminoacyl-tRNA synthetases, encoded by a domain that is distinct from the domain for aminoacylation. If they are not cleared, genetic code ambiguity is introduced (that is, a given codon in the messenger RNA will specify incorporation of more than one amino acid, resulting in the production of ‘statistical polypeptides’)^{3–6}. Here we report that an editing defect in a single tRNA synthetase in the mouse results in neurodegeneration associated with protein characteristics consistent with heterogeneous polypeptide production (Supplementary Fig. 1). These results provide a novel mechanism for the generation of misfolded proteins, which are associated with human diseases—many of which have neuron loss^{7–9}.

Purkinje cell loss in sticky mutant mice

The sticky (*sti*) mutation was identified by the rough, unkempt ‘sticky’ appearance of fur in mice homozygous for the mutation. As mice age, this rough coat is accompanied by follicular dystrophy and patchy hair loss (Supplementary Fig. 2). However, at six weeks of age mild tremors appear, which progress to overt ataxia (Supplementary Fig. 3). Histological analysis reveals a loss of cerebellar Purkinje cells in C57BL/6J.stock^{sti/sti} (*sti/sti*) mice beginning at three weeks of age (Fig. 1a–g). By six weeks, extensive Purkinje cell loss is observed,

particularly in the rostral cerebellum. This degeneration is slowly progressive so that most Purkinje cells degenerate over the course of a year, excluding the majority of these neurons in the caudally located lobule X.

To investigate the nature of Purkinje cell loss, we performed immunohistochemical studies with the apoptotic markers cleaved caspase 3 and cleaved poly(ADP-ribose) polymerase (PARP). Purkinje cells that were positive for these markers were observed in mutant, but not wild-type, animals at four weeks of age (Fig. 1h–j and data not shown). Further confirmation that these cells were undergoing apoptosis was obtained from TUNEL (TdT-mediated dUTP nick end labelling)-positive mutant Purkinje cells (Fig. 1k–m).

The *sti* mutation disrupts the *Aars* gene

A genome scan using polymorphic microsatellite markers on affected F₂ animals initially localized the *sti* mutation to chromosome 8. Further fine mapping demonstrated that *sti* resided in a 1.54-centimorgan (cM) region (45 recombinants, 2,932 meioses; Fig. 2a and Supplementary Fig. 4). Expression analysis of genes in the *sti* critical region failed to reveal any differences between wild-type and mutant cerebellar RNA. However, sequencing of complementary DNA revealed a C-to-A nucleotide change in the alanyl-tRNA synthetase gene (*Aars*) at nucleotide 2,201 of the transcript in mutant mice—predicted to cause an alanine to glutamic acid substitution at amino acid 734, a residue that is evolutionarily conserved (Supplementary Fig. 5).

To confirm that the mutation in the alanyl-tRNA synthetase protein (AlaRS) underlies Purkinje cell degeneration in *sti* mutant mice, we performed transgenic rescue experiments using a wild-type genomic DNA fragment containing the AlaRS coding region, and 7.3 kb and 1.7 kb of 5′ and 3′ flanking regions, respectively (Fig. 2b). We identified two founder lines with cerebellar expression of *Aars* at approximately five times that observed in wild-type animals (data not shown). Mutant *sti/sti* mice carrying the *Aars* transgene did not

¹The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA. ²The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ³Howard Hughes Medical Institute, Bar Harbor, Maine 04609, USA. †Present address: Department of Bioelectronics and Bioinformatics, Korea Bio Polytechnic College, Chungnam 320-905, Korea.

develop ataxia, and immunohistochemistry with calbindin-D28 demonstrated that the *sti*-associated degeneration of Purkinje cells was rescued by the transgene (Fig. 2c–f). These data demonstrate that wild-type AlaRS is necessary for Purkinje cell survival in the adult cerebellum.

Sensitivity to non-cognate amino acids

As shown in Fig. 3a, mouse AlaRS is 968 amino acids in length with a large amino-terminal catalytic domain (amino acids 1–561) for amino-acid activation and tRNA aminoacylation, and an evolutionarily conserved editing domain for hydrolytic editing of non-cognate amino acids (glycine or serine) from misaminoacylated tRNA^{Ala} (refs 10, 11). Analysis of the primary structure of AlaRS suggested that Ala734 lies within the putative editing domain (Fig. 3a). The area encompassing this residue (Fig. 3a, b (shown in purple)) has been shown to enhance aminoacylation efficiency but has no role in amino-acid activation in *Escherichia coli*¹¹. The *sti* Ala–Glu substitution seems to be within a loop region well outside of the active site of the editing domain (>15 Å) but, nevertheless, within the sequence predicted to form this domain (Fig. 3b).

In *E. coli*, mutations within the AlaRS editing domain result in increased production of misacylated Gly–tRNA^{Ala} or Ser–tRNA^{Ala}

(ref. 11). In addition to misacylation of tRNA^{Ala}, cell death occurs in editing-deficient strains grown in elevated serine or glycine¹¹. To determine whether the *sti* mutation may confer sensitivity to amino acids that may be mischarged by the mutant AlaRS enzyme, we analysed the survival of mutant and wild-type mouse embryonic fibroblasts (MEFs) challenged by the addition of individual amino acids to the medium at levels 25 × and 100 × that of unaltered media (Fig. 3c–g). Despite slightly decreased viability of both *sti/sti* and wild-type fibroblasts at increasing concentrations of amino acids, no differences in cell viability were observed following the addition of alanine, histidine or methionine ($P > 0.05$). In contrast, serine dramatically increased cell death of *sti/sti* fibroblasts in a dose-dependent manner relative to that observed in wild-type cultures ($P < 0.0001$). Mutant fibroblasts were marginally sensitive to glycine ($2.59 \pm 1.22\%$ versus $5.30 \pm 1.19\%$ cell death in wild-type versus mutant cells; values are mean $\pm$ s.e.m.). The specific sensitivity to serine is consistent with defective editing in the *sti* mutant AlaRS protein. Also, *sti/+* fibroblasts show an increased sensitivity to serine that is intermediate to that of wild-type and *sti/sti* fibroblasts (Fig. 3h). This contrasts with the normal phenotype of *sti/+* mice, suggesting that environmental challenges can result in gene dosage effects of the *sti* mutation.

Defects in editing of non-cognate amino acids from mischarged tRNAs should lead to misincorporation of these amino acids during protein synthesis, which would in turn lead to the accumulation of misfolded/unfolded proteins. To determine if serine-induced cell death in *sti/sti* cells is correlated with the accumulation of unfolded proteins, we analysed ubiquitination of proteins in serine-treated fibroblasts pretreated with mitomycin C to inhibit cell division and dilution of misfolded proteins (Fig. 3i). In the absence of serine addition, polyubiquitinated proteins were greatly increased in *sti/sti* fibroblasts, indicating that misfolded/unfolded proteins accumulate in these cells. Ubiquitinated proteins increased in both mutant and wild-type cells on addition of serine to the media, indicating that high concentrations of serine cause protein misfolding even in wild-type cells, perhaps by interfering with AlaRS editing. Consistent with misfolded protein elevation, levels of stress-inducible, cytosolic heat shock protein 72 (HSP72) were increased in *sti/sti* fibroblasts, and also increased on the addition of serine to wild-type cells (Fig. 3i).

The *sti* mutation disrupts AlaRS editing

For functional analysis, we introduced the A734E mutation into both the mouse and the human enzyme, which is 92% identical (including Ala734) to mouse AlaRS. Far-ultraviolet circular dichroism (far-UV CD) spectral analysis of purified wild-type and mutant proteins had identical minima and indistinguishable far-UV CD spectra, suggesting that the A734E substitution does not induce major changes in local secondary structure of AlaRS (Supplementary Methods and Supplementary Fig. 6).

A single G3:U70 acceptor stem base pair marks tRNA^{Ala} throughout evolution for charging with alanine. As a consequence, AlaRS proteins from bacteria to humans charge tRNA^{Ala} across species^{12,13}. We analysed the effect of the A734E substitution on *in vitro* aminoacylation activity of human and mouse enzymes on *E. coli* tRNA^{AlaGAC} or human tRNA^{AlaUGC}, respectively, both of which retain the G3:U70 base pair in the acceptor stem that is crucial for efficient AlaRS recognition¹⁴. Wild-type and mutant A734E human AlaRS purified from *E. coli* showed the same kinetics for aminoacylation for tRNA^{Ala} (wild type: $k_{cat} = 1.5 \text{ s}^{-1}$ and $K_m = 5.7 \mu\text{M}$; mutant: $k_{cat} = 1.6 \text{ s}^{-1}$ and $K_m = 6.8 \mu\text{M}$). Similarly, no differences in aminoacylation kinetics were observed between recombinant wild-type ($k_{cat} = 1.5 \text{ s}^{-1}$ and $K_m = 0.49 \mu\text{M}$) and mutant ($k_{cat} = 1.5 \text{ s}^{-1}$ and $K_m = 0.73 \mu\text{M}$) mouse enzyme (Supplementary Fig. 7).

To test the effect of the *sti* mutation on editing, we assayed deacylation of mischarged tRNA^{Ala} with mutant and wild-type enzyme. The viability of *sti* embryonic fibroblasts is more severely

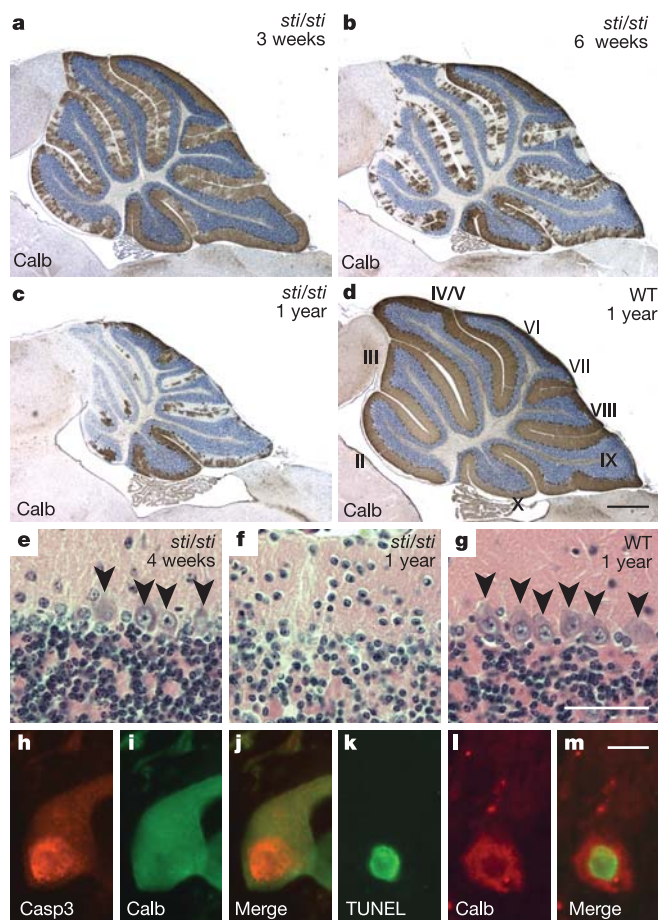


Figure 1 | Sticky mutant pathology. **a–d**, Calbindin D-28 (Calb) immunohistochemistry of sagittal sections of cerebella from 3-week-old (**a**), 6-week-old (**b**) or 12-month-old (**c**) *sti/sti* mutant and 12-month-old wild-type (WT) (**d**) mice. Cerebellar lobules are indicated by Roman numerals. **e–g**, Haematoxylin and eosin staining of Purkinje cells (arrowheads) in lobule II of cerebella from 1-month-old (**e**) or 12-month-old (**f**) *sti/sti* mutant and 12-month-old wild-type (**g**) mice. **h–m**, Cleaved caspase 3 (Casp3) immunohistochemistry (**h–j**) and TUNEL analysis (**k–m**) of 4-week-old mutant cerebella. Scale bars: **a–d**, 500 μm ; **e–g**, 50 μm ; **h–m**, 10 μm .

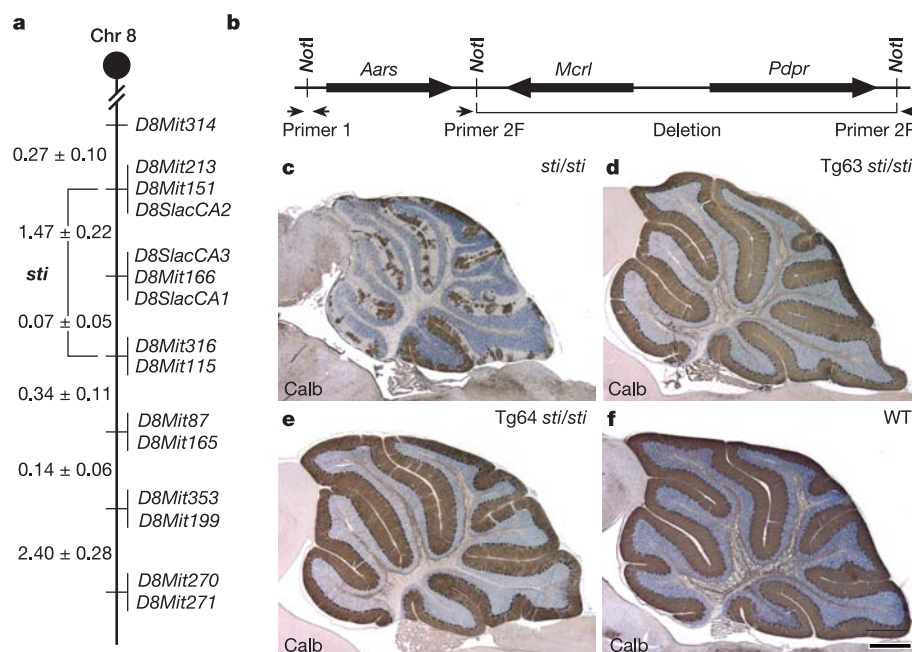


Figure 2 | The *sti* mutation is identified in the alanyl-tRNA synthetase (*Aars*) gene. **a**, The *sti* mutation was mapped to chromosome 8 (shown in cM $\pm$ s.e.m.). **b**, The RP24-359N5 bacterial artificial chromosome (BAC) contains the *Aars*, *Mrc1* and *Pdpr* genes, of which the latter two were deleted by partial digestion with *NotI*. Positions of primers used to genotype

transgenic mice are shown. **c–f**, Immunohistochemistry of sagittal sections of cerebella from 3-month-old *sti/sti* (**c**), transgenic (Tg) *sti/sti* carrying the *Aars* transgene from line 63 or line 64 (**d**, **e**, respectively), and wild-type (**f**) mice using the calbindin D-28 antibody. Scale bar: 500 μ m.

affected when challenged with serine than with glycine. Thus, we first examined the effect this mutation had on the ability of the enzyme to hydrolyse human Ser-tRNA^{Ala}. With the mutant enzyme, deacylation of Ser-tRNA^{Ala} was diminished by a reproducible 40–50% (Fig. 4a). In contrast, no deacylation of Ala-tRNA^{Ala} above background was observed with either mutant or wild-type enzyme, excluding the possibility of a loss in editing specificity by the *sti* mutant (Supplementary Fig. 8).

A loss of editing activity should result in the production and release of misacylated tRNAs. *In vivo*, these incorrect products are

captured by elongation factors, delivered to the ribosome, and incorporated into nascent peptides. In the presence of the elongation factor (EF) EF-Tu to capture and sequester misacylated tRNA^{Ala}, a significantly higher accumulation of Ser-tRNA^{Ala} was observed in reactions using the mutant versus wild-type human enzyme (Fig. 4b). Although the deacylation of Ser-tRNA^{Ala} was nearly indistinguishable between the mutant and wild-type mouse enzymes, the mutant mouse enzyme also produced increased Ser-tRNA^{Ala} (Fig. 4c, d). Notably, although mouse AlaRS produces significant quantities of Ser-tRNA^{Ala}, it fails to generate Gly-tRNA^{Ala} (Supplementary Fig. 9),

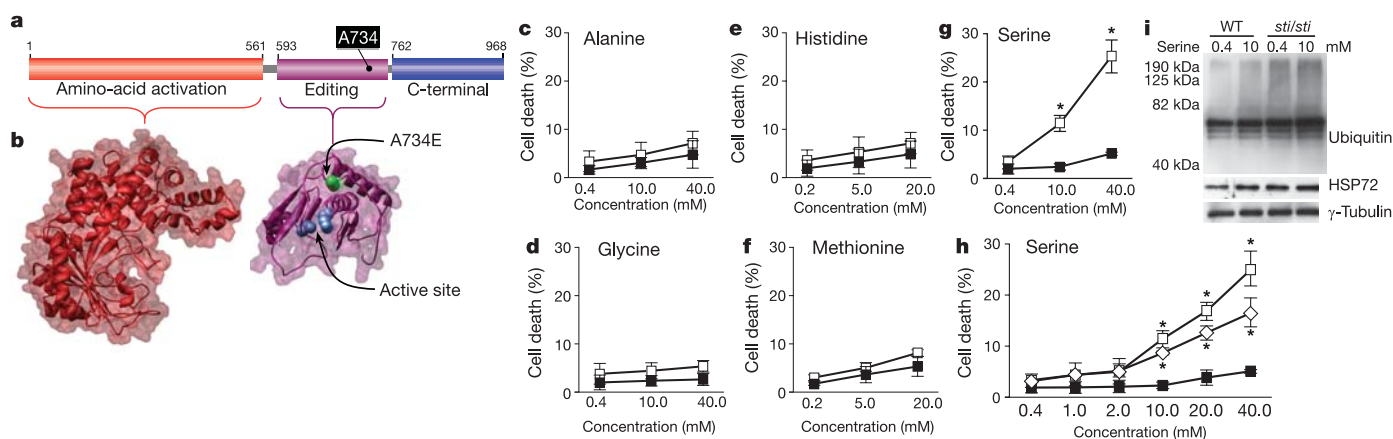


Figure 3 | Mutant fibroblasts are selectively sensitive to serine. **a, b**, AlaRS domains (**a**) and their three-dimensional structures (**b**). The amino-acid activation domain (red) and editing domain (purple) is modelled on *Aquifex aeolicus* AlaRS³⁴ and the *Pyrococcus horikoshii* free-standing AlaRS editing domain homologue²³, respectively. Blue spheres show the location of crucial active site residues. Ala 734 is depicted as a green sphere. **c–g**, Wild-type (black boxes) and mutant (open boxes) MEFs were treated with increasing concentrations of alanine (**c**), glycine (**d**), histidine (**e**), methionine (**f**) and

serine (**g**). **h**, Viability curves for wild-type (black boxes), *sti*/+ (open diamonds) and *sti/sti* (open boxes) MEFs in the presence of increasing concentrations of serine. Values are the means of three or more independent experiments $\pm$ s.e.m. *, $P < 0.0001$. **i**, Western blot of extracts from mitomycin-C-arrested mutant and wild-type cells grown in unsupplemented (0.4 mM) or serine-supplemented (10 mM) media. Anti- γ -tubulin antibody was used as a loading control.

consistent with differential sensitivity of *sti/sti* embryonic fibroblasts to serine over glycine.

In vivo effects of decreased editing

The partial loss of editing function of A734E AlaRS and the accompanying increase in misacylated tRNA^{Ala} would be predicted to result in the increased production of proteins with misincorporated amino acids, probably leading to the accumulation of misfolded proteins. Electron microscopy analysis of *sti/sti* cerebella revealed multilamellar membranous structures enveloping the cytoplasm and organelles, characteristic of autophagosomes (Fig. 5a–c). In addition, electron-dense globular structures, reminiscent of protein inclusions, were observed in the cytoplasm and perinuclear region of mutant Purkinje cells.

Ubiquitinated proteins are commonly found in neuronal inclusions in many neurodegenerative disorders¹⁵. Immunofluorescence for ubiquitin demonstrated intense punctate staining throughout the cytoplasm of *sti/sti* Purkinje cells that was not seen in the wild type (Fig. 5d–g). These puncta were also observed in the axons and dendrites of many Purkinje cells (data not shown). In addition, ubiquitin immunoreactivity was observed in the nucleolus of mutant Purkinje cells, suggesting that nucleolar proteins are also misfolded (Fig. 5h, i).

An increase in misfolded/unfolded proteins can result in the induction of molecular chaperones that bind to interactive surfaces of abnormal proteins and facilitate refolding or degradation of misfolded polypeptides^{16–18}. Both cytoplasmic and endoplasmic-reticulum-localized members of the HSP70 chaperone family are associated with neuronal protein aggregates in many progressive neurodegenerative disorders¹⁹. Immunofluorescence for inducible, cytosolic HSP72 and constitutive, cytosolic heat shock cognate 70 (HSC70) demonstrated that these chaperones were induced in *sti* mutant Purkinje cells that also showed an increase in ubiquitinated

proteins (Fig. 5j, k and Supplementary Fig. 10). Similarly, levels of the HSP70 co-chaperone HSP40 were markedly upregulated in these neurons (Supplementary Fig. 10).

Misfolded proteins within the endoplasmic reticulum can result in an imbalance between the unfolded protein load and the capacity of the endoplasmic reticulum protein folding machinery, also known as endoplasmic reticulum stress²⁰. The intrinsic response to restore endoplasmic reticulum homeostasis is known as the unfolded protein response (UPR). To test whether the accumulation of unfolded proteins in *sti/sti* Purkinje cells results in endoplasmic reticulum stress, we examined the expression of the endoplasmic reticulum chaperone BiP (also known as GRP78 or HSPA5), and an endoplasmic-reticulum-stress-induced transcription factor CHOP (also known as GADD153 or DDIT3). Immunofluorescence analysis demonstrated that BiP is transiently upregulated in many Purkinje cells of the *sti/sti* cerebellum at two weeks of age compared with wild-type Purkinje cells where levels are very low (Supplementary Fig. 10). CHOP expression is apparent in mutant Purkinje cells, but not wild-type cells, by three weeks of age (Fig. 5l–o). It remains upregulated at five weeks of age, a time at which many Purkinje cells degenerate, indicating a potential role of CHOP in apoptosis, as previously suggested²¹.

Discussion

Small amounts of misfolded proteins with abnormal conformations can serve to nucleate additional proteins, forming larger, potentially toxic, protein aggregates or leading to the loss of function of aggregated proteins²². Postmitotic cells, in particular neurons, seem to be extremely sensitive to misfolded proteins, possibly because these potentially toxic species cannot be diluted by cell division. Indeed, the most dramatic phenotype in *sti* mutant mice is neuron loss. Like most neurodegenerative diseases, pathology is only observed in

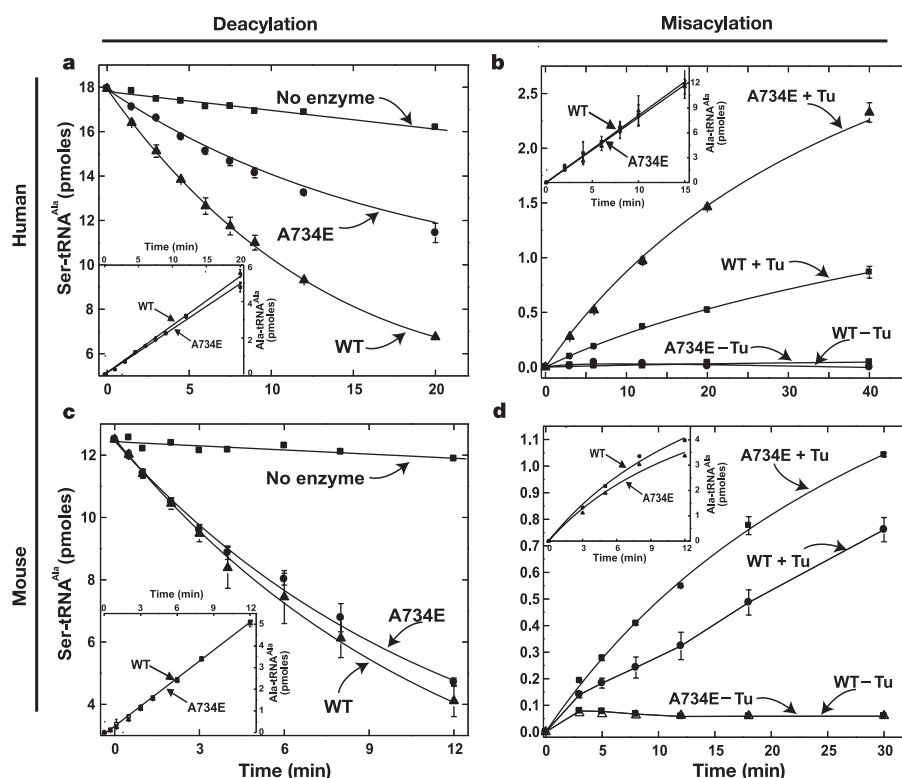


Figure 4 | Editing deficiency in *sti* mutant AlaRS. a, Deacylation of [³H]Ser-tRNA^{Ala} by wild-type or A734E human AlaRS. Inset, aminoacylation control with the same enzyme dilution used in the experiment shown in the main panel. **b**, Misacylation of tRNA^{Ala} with serine by wild-type or A734E human AlaRS in the presence of *Thermus thermophilus* EF-Tu. Inset,

aminoacylation dilution control. **c**, Deacylation of [³H]Ser-tRNA^{Ala} by mouse AlaRS. Inset, aminoacylation control with the same enzyme dilution used in the experiment shown in the main panel. **d**, Misacylation of tRNA^{Ala} by mouse AlaRS. Inset, aminoacylation dilution control. Values are the means of two independent experiments $\pm$ s.d.

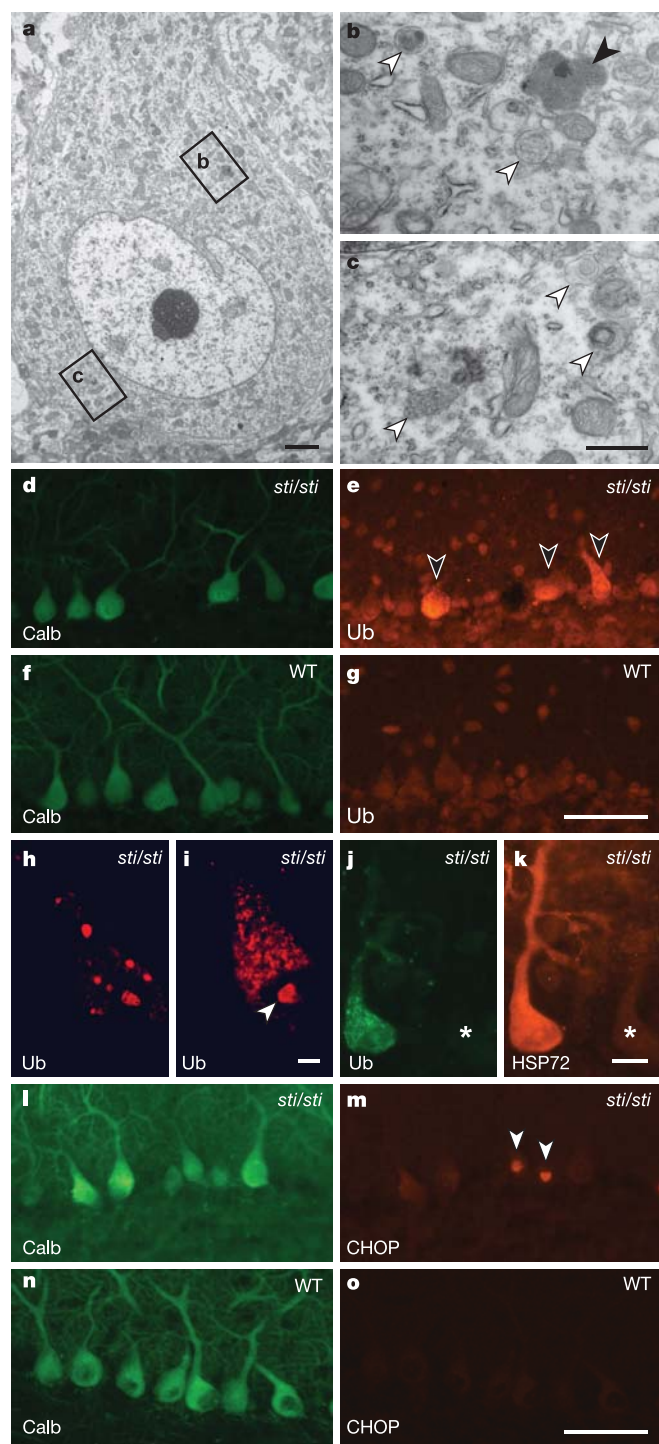


Figure 5 | Accumulation of misfolded proteins in *sti/sti* Purkinje cells. **a–c**, Electron micrographs of *sti/sti* Purkinje neurons at 3 weeks of age. Boxed areas in **a** are shown in detail in **b** and **c**. Autophagosome-like vacuoles (open arrowheads) and globular structures reminiscent of protein accumulations (closed arrowhead), are indicated. **d–g**, Immunofluorescence with antibodies to calbindin D-28 or ubiquitin (Ub) on 3-week-old wild-type or *sti/sti* cerebella. Arrowheads indicate punctate ubiquitin staining in the cytoplasm of mutant Purkinje cells. **h, i**, Confocal microscopy analysis of *sti/sti* Purkinje cells stained with ubiquitin antibodies demonstrates large puncta of staining throughout the cytoplasm and nucleolus (arrowhead). **j, k**, HSP72 staining is induced in cells with punctate ubiquitin staining. An asterisk indicates an adjacent Purkinje cell without punctate ubiquitin staining or induced expression of HSP72. **l–o**, Immunofluorescence of 3-week-old *sti/sti* (**l, m**) and wild-type (**n, o**) cerebella with CHOP (arrowheads) and calbindin D-28 antibodies. Scale bars: **a**, 2 μ m; **b, c**, 500 nm; **d–o**, 50 μ m.

specific neurons in *sti* mutant mice: in this case, the cerebellar Purkinje cells. The cause of this specificity is not clear. In the human and mouse, cytoplasmic AlaRS, like most tRNA synthetases, is encoded by a single gene, and a separate gene encodes the mitochondrial form. However, the mouse genome does contain AlaRS-like genes that encode proteins with free-standing tRNA synthetase editing domains. Although these proteins have been shown to edit mischarged tRNAs in prokaryotes^{23,24}, it is not known whether these function in a similar manner in eukaryotes, potentially compensating for the loss of AlaRS-mediated editing in other neuronal populations. Alternatively, Purkinje cells might be inherently less efficient at degrading misfolded proteins or more sensitive to the deleterious effects of these polypeptides, as suggested by the many human repeat-expansion diseases that are associated with ataxia and Purkinje cell loss²⁵. Furthermore, recent studies in mouse and human demonstrate that the loss of the widely expressed co-chaperone of the endoplasmic reticulum chaperone BiP, SIL1, specifically results in misfolded proteins and subsequent degeneration of Purkinje cells^{26–28}.

Our data demonstrate the importance of proofreading of non-cognate amino acids by aminoacyl synthetases in terminally differentiated neurons. Although our results are most consistent with loss of editing activity, we cannot completely rule out the possibility that the *sti* mutation disrupts a novel function of AlaRS. However, our independent experiments have shown that inducible expression of an editing-defective valyl-tRNA synthetase in cultured mammalian cells leads to mistranslation and apoptosis, which is exacerbated by the addition to the culture of a non-canonical amino acid³⁵. Furthermore, because editing domains of tRNA synthetases are functionally independent of those for the essential aminoacylation function, mild defects in editing could be transmitted from generation to generation without disruption of protein synthesis, raising the possibility that some heritable diseases are connected to mild mutations in tRNA synthetase editing functions that, in turn, generate misfolded proteins.

METHODS

Mice. The sticky mutation arose spontaneously in an unknown stock mouse and was transferred to a C57BL/6J background. The truncated RP24-359N5 BAC was microinjected into the pronuclei of C57BL/6J embryos. Transgenic mice were identified as described in the Supplementary Methods. An intersubspecific intercross (C57BL/6J.stock^{sti/sti} × CAST/Ei) was used for genetic mapping.

Immunohistochemistry and electron microscopy. Electron microscopy studies, TUNEL assays, immunohistochemistry and immunofluorescence were performed as described previously^{28,29}, or using antibodies to HSC70, HSP72 and HSP40 (Stressgen Bioreagents; 1:200 dilution).

Mouse embryonic fibroblast cultures. MEFs were prepared by standard protocols³⁰. Amino acids were added to the culture media for 24 h; cells were collected and stained with propidium iodide. The percentage of cell death was measured by fluorescence-activated cell sorting (FACS) analysis. Statistical significance was determined by analysis of covariance (ANCOVA) using SAS software (SAS Institute).

Western blots. MEFs were treated with mitomycin C (1 μ g ml⁻¹) for 2 h, before a 24-h incubation in serine-supplemented media. Blots were incubated with antibodies to ubiquitin (Cell Signalling; 1:1,000), HSP72 (Stressgen Bioreagents; 1:1,000) and γ -tubulin (Sigma; 1:2,000).

Aminoacylation, deacylation and misacylation assays. Overexpression and purification of *E. coli* tRNA^{AlaGCG} and human tRNA^{AlaUGC} was accomplished as described³¹. Purified human tRNA^{AlaUGC} (34 μ M) was aminoacylated by *E. coli* AlaRS (C666A/Q584H) as described¹¹. Deacylation assays were performed at pH 7.5, 26 °C as described^{11,32}, using human or mouse (8 nM and 2 nM, respectively) wild-type mutant A734E AlaRS, and 6.2 μ M mischarged tRNA^{Ala}. Misacylation assays were performed at pH 7.5, 26 °C using human or mouse (1.3 μ M and 3.7 μ M, respectively) wild-type or mutant A734E AlaRS, 10 μ M serine (or 8.8 μ M glycine), 10 μ M, tRNA^{Ala}, 4 mM ATP, 4 μ M *Thermus thermophilus* EF-Tu, and 4 ng μ l⁻¹ of inorganic pyrophosphatase (Roche) in EF-Tu buffer^{11,33}. EF-Tu was activated as described³³.

Received 3 February; accepted 19 July 2006.

Published online 13 August 2006.

1. Loftfield, R. B. & Vanderjagt, D. The frequency of errors in protein biosynthesis. *Biochem. J.* 128, 1353–1356 (1972).

2. Jakubowski, H. & Goldman, E. Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* **56**, 412–429 (1992).
3. Bacher, J. M., de Crécy-Lagard, V. & Schimmel, P. R. Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. *Proc. Natl Acad. Sci. USA* **102**, 1697–1701 (2005).
4. Doring, V. *et al.* Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* **292**, 501–504 (2001).
5. Nangle, L. A., de Crécy-Lagard, V., Doring, V. & Schimmel, P. Genetic code ambiguity. Cell viability related to the severity of editing defects in mutant tRNA synthetases. *J. Biol. Chem.* **277**, 45729–45733 (2002).
6. Pezo, V. *et al.* Artificially ambiguous genetic code confers growth yield advantage. *Proc. Natl Acad. Sci. USA* **101**, 8593–8597 (2004).
7. Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).
8. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
9. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nature Med.* **10** (Suppl.), S10–S17 (2004).
10. Dock-Bregeon, A. *et al.* Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. *Cell* **103**, 877–884 (2000).
11. Beebe, K., Ribas De Pouplana, L. & Schimmel, P. Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. *EMBO J.* **22**, 668–675 (2003).
12. Hou, Y. M. & Schimmel, P. Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. *Biochemistry* **28**, 6800–6804 (1989).
13. Ripmaster, T. L., Shiba, K. & Schimmel, P. Wide cross-species aminoacyl-tRNA synthetase replacement *in vivo*: yeast cytoplasmic alanine enzyme replaced by human polymyositis serum antigen. *Proc. Natl Acad. Sci. USA* **92**, 4932–4936 (1995).
14. Hou, Y. M. & Schimmel, P. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**, 140–145 (1988).
15. Ross, C. A. & Pickart, C. M. The ubiquitin–proteasome pathway in Parkinson's disease and other neurodegenerative diseases. *Trends Cell Biol.* **14**, 703–711 (2004).
16. Welch, W. J. Role of quality control pathways in human diseases involving protein misfolding. *Semin. Cell Dev. Biol.* **15**, 31–38 (2004).
17. Dobson, C. M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15**, 3–16 (2004).
18. Barral, J. M., Broadley, S. A., Schaffar, G. & Hartl, F. U. Roles of molecular chaperones in protein misfolding diseases. *Semin. Cell Dev. Biol.* **15**, 17–29 (2004).
19. Muchowski, P. J. & Wacker, J. L. Modulation of neurodegeneration by molecular chaperones. *Nature Rev. Neurosci.* **6**, 11–22 (2005).
20. Xu, C., Bailly-Maitre, B. & Reed, J. C. Endoplasmic reticulum stress: cell life and death decisions. *J. Clin. Invest.* **115**, 2656–2664 (2005).
21. Zinszner, H. *et al.* CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. *Genes Dev.* **12**, 982–995 (1998).
22. Murphy, R. M. Peptide aggregation in neurodegenerative disease. *Annu. Rev. Biomed. Eng.* **4**, 155–174 (2002).
23. Sokabe, M., Okada, A., Yao, M., Nakashima, T. & Tanaka, I. Molecular basis of alanine discrimination in editing site. *Proc. Natl Acad. Sci. USA* **102**, 11669–11674 (2005).
24. Ahel, I., Korencic, D., Ibba, M. & Soll, D. Trans-editing of mischarged tRNAs. *Proc. Natl Acad. Sci. USA* **100**, 15422–15427 (2003).
25. Gatchel, J. R. & Zoghbi, H. Y. Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Rev. Genet.* **6**, 743–755 (2005).
26. Senderek, J. *et al.* Mutations in *SIL1* cause Marinesco–Sjogren syndrome, a cerebellar ataxia with cataract and myopathy. *Nature Genet.* **37**, 1312–1314 (2005).
27. Anttonen, A. K. *et al.* The gene disrupted in Marinesco–Sjogren syndrome encodes SIL1, an HSPA5 cochaperone. *Nature Genet.* **37**, 1309–1311 (2005).
28. Zhao, L., Longo-Guess, C., Harris, B. S., Lee, J. W. & Ackerman, S. L. Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. *Nature Genet.* **37**, 974–979 (2005).
29. Ackerman, S. L. *et al.* The mouse rostral cerebellar malformation gene encodes an UNC-5-like protein. *Nature* **386**, 838–842 (1997).
30. Hogan, B. *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1986).
31. Xue, H., Shen, W. & Wong, J. T. Purification of hyperexpressed *Bacillus subtilis* tRNA^{Trp} cloned in *Escherichia coli*. *J. Chromatogr.* **613**, 247–255 (1993).
32. Beebe, K., Merriman, E., Ribas De Pouplana, L. & Schimmel, P. A domain for editing by an archaeobacterial tRNA synthetase. *Proc. Natl Acad. Sci. USA* **101**, 5958–5963 (2004).
33. Pleiss, J. A. & Uhlenbeck, O. C. Identification of thermodynamically relevant interactions between EF-Tu and backbone elements of tRNA. *J. Mol. Biol.* **308**, 895–905 (2001).
34. Swairjo, M. A. *et al.* Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Mol. Cell* **13**, 829–841 (2004).
35. Nangle, L. A., Motta, C. M. & Schimmel, P. Global effects of mistranslation from an editing defect in mammalian cells. *Chem. Biol.* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank The Jackson Laboratory sequencing, microchemistry, histology, bioimaging and microinjection services for their contributions. We also thank L. Dionne and K. Seburn for rotorod testing and data analysis, J. Szatkiewicz for assistance with statistical analysis, J. Torrance for graphics assistance, P. O'Maille, W. Waas and A. Wolfson for gifts of plasmids, and C. Motta and E. Merriman for facilitating the expression and purification of proteins. We are also grateful to the laboratory of J. Kelly for CD spectrometer assistance and to R. Burgess and P. Nishina for comments on the manuscript. This work was supported by grants from the National Institute of Neurological Disorders and Stroke and the National Institute on Aging to S.L.A., the National Institute of General Medical Sciences and a fellowship from the National Foundation for Cancer Research to P.S., a grant from the National Center for Research Resources to M.T.D., and an institutional National Cancer Institute core grant (JAX). S.L.A. is an investigator of the Howard Hughes Medical Institute.

Author Contributions J.W.L. designed and performed mouse and cell culture experiments, K.B. and L.A.N. designed and performed biochemical analyses, J.J. performed mutation analysis, M.T.D. provided the congenic *sti* mice and oversaw initial mapping experiments, S.A.C. and C.M.L.-G. performed genetic mapping experiments, J.P.S. performed hair pathological analysis, P.S. and S.L.A. designed and supervised experiments. All authors discussed the results and commented on the manuscript, which was written by J.W.L., K.B., P.S. and S.L.A.

Author Information The sequence for mouse *Aars* has been deposited in GenBank under the accession number AY223875; sequence-tagged sites (STSs) for *D8SlacA1* (DQ386090), *D8SlacA2* (DQ386088) and *D8SlacA3* (DQ386089) can also be found in GenBank. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L.A. (susan.ackerman@jax.org).

LETTERS

Developing space weathering on the asteroid 25143 Itokawa

Takahiro Hiroi¹, Masanao Abe², Kohei Kitazato^{2,3}, Shinsuke Abe⁴, Beth E. Clark⁵, Sho Sasaki⁶, Masateru Ishiguro⁷ & Olivier S. Barnouin-Jha⁸

Puzzlingly, the parent bodies of ordinary chondrites (the most abundant type of meteorites) do not seem to be abundant among asteroids. One possible explanation is that surfaces of the parent bodies become optically altered, to become the S-type asteroids which are abundant in the main asteroid belt. The process is called 'space weathering'—it makes the visible and near-infrared reflectance spectrum of a body darker and redder¹. A recent survey of small, near-Earth asteroids suggests that the surfaces of small S asteroids may have developing stages of space weathering². Here we report that a dark region on a small (550-metre) asteroid—25143 Itokawa—is significantly more space-weathered than a nearby bright region. Spectra of both regions are consistent with those of LL5-6 chondrites after continuum removal³. A simple calculation⁴ suggests that the dark area has a shorter mean optical path length and about 0.04 per cent by volume more nanophase metallic iron particles than the bright area. This clearly shows that space-weathered materials accumulate on small asteroids, which are likely to be the parent bodies of LL chondrites. We conclude that, because LL meteorites are the least abundant of ordinary (H, L, and LL) chondrites, there must be many asteroids with ordinary-chondrite compositions in near-Earth orbits.

The Japanese Hayabusa spacecraft was launched on 9 May 2003 and made a successful rendezvous with the asteroid 25143 Itokawa on 12 September 2005. During its three-month rendezvous and sampling period until early December 2005, the Near-Infrared Spectrometer³ (NIRS) onboard Hayabusa took continuous point-target near-infrared reflectance spectra covering an effective wavelength range of $\lambda = 0.76\text{--}2.1\ \mu\text{m}$. Although the footprint positions for the majority of NIRS observations are not accurately located owing to failures of two reaction wheels of the spacecraft, some positions are reasonably well known from simultaneous observations of the Asteroid Multiband Imaging Camera⁵ (AMICA), also onboard the spacecraft. AMICA images clearly show large variations in both brightness and colour, correlated with each other over the entire surface of Itokawa⁶.

Such evidence of space weathering can be addressed more precisely using NIRS data. A preliminary study³ showed that the surface composition of Itokawa was probably the LL5 or LL6 chondrite meteorite. To build on this study³, here we focused on the space weathering on Itokawa near locations Tsukuba and Sagami-hara in NIRS observations on 25 October 2005 (shown in Fig. 1). Figure 2 shows variations in reflectance and the $1\text{-}\mu\text{m}$ -band continuum slope in the time sequence during the above NIRS observations. There is good correlation between reflectance and the continuum slope: the darker the reflectance, the redder the continuum. Four data points,

two from each of the areas (bright and blue, and dark and red) near Tsukuba and Sagami-hara were chosen for analyses.

Reflectance spectra of the above chosen points are photometrically corrected³ to 30° incidence and 0° emergence angles and averaged

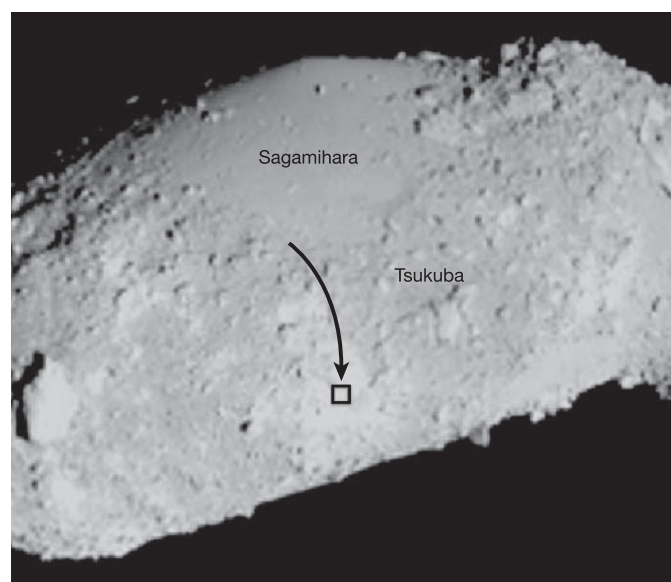


Figure 1 | A monochromatic image of asteroid Itokawa with NIRS observation points indicated. This AMICA v-band image (time ID: 2489416612) showing Sagami-hara and Tsukuba was taken at 8:16:21 UT on 25 October 2005 at a solar phase angle of about 21° and a distance from the asteroid of 5.3 km. The NIRS footprint—about 6.6 m in size¹² when this image was taken—is indicated as an open square, and its movement as an arrow. The location of the NIRS footprint requires very accurate spacecraft location information. This was achieved in a non-traditional manner by using data on the illuminated centre of figure XY of the asteroid obtained by the Wide Angle Camera (WAC) aboard the Hayabusa spacecraft. Using knowledge of the alignment between NIRS, the Hayabusa laser altimeter and this WAC imager, we were able to use the XY information, the laser altimeter ranges, and the most recent shape model of Itokawa¹³ to determine the spacecraft location at 2-min intervals. From the spacecraft attitude information and a simple spacecraft trajectory model, we determined the location of the spacecraft within 10 m at every 1-s interval each time the laser altimeter ranged to the surface of Itokawa. With this spacecraft location we were able to identify the location of the NIRS footprint in most instances to within 10 m.

¹Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. ²Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagami-hara, Kanagawa 229-8510, Japan. ³Department of Earth and Planetary Science, University of Tokyo, Tokyo 113-0033, Japan. ⁴Graduate School of Science and Technology, Kobe University, Kobe 657-8501, Japan. ⁵Ithaca College, 267 Center for Natural Sciences, Ithaca, New York 14850-7288, USA. ⁶RISE Project Office, National Astronomical Observatory of Japan, Mizusawa, Oshu 023-0861, Japan. ⁷School of Earth Environmental Sciences, College of Natural Sciences, Seoul National University, Seoul 151-742, Korea. ⁸Johns Hopkins University Applied Physics Laboratory, Johns Hopkins Road, Laurel, Maryland 20723-6099, USA.

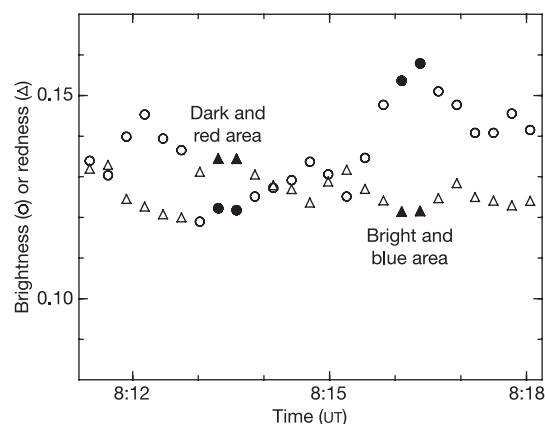


Figure 2 | Variations in brightness and redness of NIRS spectra near Tsukuba and Sagami. NIRS at Itokawa on 25 October 2005. The brightness is defined as reflectance at 0.76 μm (open and filled circles) and the redness as the 1- μm band continuum slope (1.54 $\mu\text{m}/0.76 \mu\text{m}$) (open and filled triangles). Continuum slopes are plotted by scaling with a factor of 0.1. The NIRS observations were performed at intervals of 17 s. Data points plotted in filled triangles and circles indicate spectra which are darker than 12.3% and redder than 1.33 (dark and red area), or brighter than 15.2% and bluer than 1.23 (bright and blue area) used for spectral analyses.

over each region. The spectra are plotted in Fig. 3 and compared with an LL5 chondrite spectrum after continuum removal. Consistent with earlier studies^{3,7}, these NIRS spectra exhibit similar band features around 1.0, 1.25 and 1.95 μm to those of this LL5 chondrite. After continuum removal, all three spectra look very similar to one another except for the depths of absorption bands. The mismatch around 1.7 μm in wavelength is due to the known calibration error. This suggests that they are made of the same material (similar to LL5 or LL6 chondrite), with different degrees of space weathering and/or mean optical path length (MOPL). The MOPL can be the effective grain size if the material consists of particulate, monomineralic grains, but the MOPL of a rock may depend not only the mineral grain sizes but also physical conditions such as porosity.

To estimate the degrees of space weathering in these two regions, we used Hapke's space weathering model⁴. He proposed a formula for the absorption coefficient $\alpha(\lambda)$ of a material containing nano-phase metallic iron (npFe⁰) particles:

$$\alpha(\lambda) = \alpha_h(\lambda) + \frac{36\pi}{\lambda} \phi z(\lambda) \quad (1)$$

where

$$z(\lambda) = \frac{n_h^3 n_{\text{Fe}} k_{\text{Fe}}}{(n_{\text{Fe}}^2 - k_{\text{Fe}}^2 + 2n_h^2)^2 + (2n_{\text{Fe}} k_{\text{Fe}})^2} \quad (2)$$

where α_h is the absorption coefficient of the host material, λ is the wavelength of light, n_h is the real part of the refractive index of the host material, n_{Fe} and k_{Fe} are the real and imaginary parts of the refractive index of iron, and ϕ is the volume fraction of npFe⁰ particles in the host material. If we assume that the absorbance spectrum of the Alta'ameem sample and the surface material of Itokawa can be approximated by the negative natural logarithm of its reflectance spectrum, we can obtain, from equation (1):

$$\ln R(\lambda) \equiv - \left\{ \alpha_h(\lambda) + \frac{36\pi}{\lambda} \phi z(\lambda) \right\} d_e \quad (3)$$

where $R(\lambda)$ denotes reflectance at λ , and d_e denotes the MOPL.

The absorption coefficient spectrum of the Alta'ameem sample (<125 μm) can be obtained using equation (3) with no npFe⁰ ($\phi = 0$) from its laboratory reflectance spectrum: for example, by assuming its effective grain size to be the median 62.5 μm . Assuming the host material of Itokawa's surface has the same optical properties

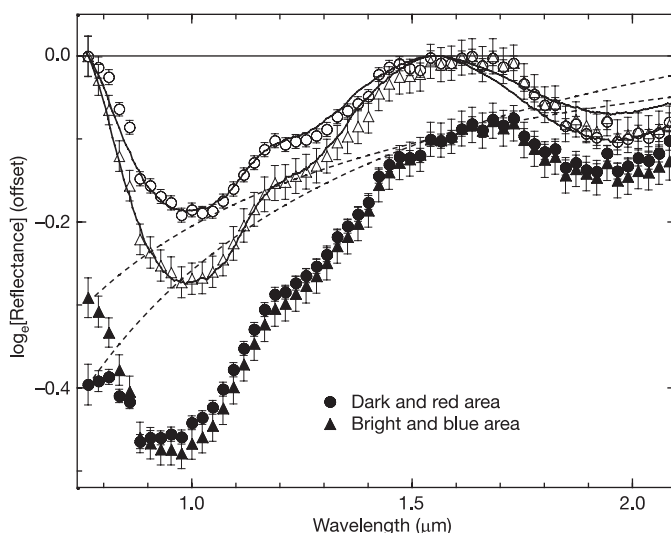


Figure 3 | Comparison of reflectance spectra of two areas on Itokawa with an ordinary chondrite spectrum. NIRS at Itokawa on 25 October 2005. Plotted here are natural logarithm of average reflectance spectra of the dark and red area and the bright and blue area (filled markers), their continuum-removed spectra (open markers), and the continuum-removed and scaled spectrum of Alta'ameem (LL5) chondrite¹⁴ powder sample (<125 μm). Error bars are 1σ . Both the NIRS spectra and the Alta'ameem spectrum are for the same viewing geometry of 30° incidence and 0° emergence angles. The continuum (broken line) is defined as a linear-to-wavenumber curve which passes the shortest-wavelength data point (at 0.76 μm) and is tangential to the spectrum around 1.55 μm , connected with a constant (flat line) longward of that contact point. The continuum-removed Alta'ameem spectrum (solid line) was scaled by factors of 0.43 and 0.63 to fit with the spectra of the dark and red area and the bright and blue area, respectively.

as does the Alta'ameem sample, we use the above-obtained absorption coefficient as $\alpha_h(\lambda)$ in equation (3). From a previous study⁸ that found that LL5-6 chondrites have mineral assemblages dominated by olivine (about 55% in modal abundance as opposed to 25% of pyroxene abundance), which has an average chemical composition of $\text{Fe}/(\text{Fe} + \text{Mg}) = \sim 0.4$ (Fa 40: fayalite 40% and forsterite 60% in the olivine solid-solution series), we can derive the real part of the refractive index of the host material n_h to be 1.76 at the visible wavelength (0.55 μm)⁹. The corresponding spectral shape is assumed to be $n_h(\lambda) = 1.6328 + 0.06998/\lambda$ (ref. 10). Then, the $z(\lambda)$ function in equation (3) can be calculated using equation (2) and by using the data on the real and imaginary parts of the refractive index of metallic iron¹¹ extrapolated to the longer wavelength, as was done in Hapke's space weathering model⁴.

We are now ready to estimate the percentage volume of npFe⁰ (ϕ) and MOPL (d_e) in equation (3) for the natural log reflectance spectra of the two areas (dark and red, and bright and blue) by optimizing these two parameters for the best fit between the measured spectrum for each region on the left hand and the model spectrum calculated on the right hand of equation (3). The results are shown in Fig. 4. As is expected, both areas require some amounts of npFe⁰ and smaller MOPL than those for the Alta'ameem sample, where the dark and red area, having 0.069 vol.% npFe⁰, is more space-weathered than the bright and blue area, having 0.031 vol.% npFe⁰. The average of these two amounts of npFe⁰ is consistent with the result of a ground-based telescopic work⁷, which estimated it as 0.05 vol.%.

We note that the natural log reflectance values used in the optimization process and shown in Fig. 4 are offset to 0 at $\lambda = 1.54 \mu\text{m}$. Owing to the surface roughness and incident angle variation, a wavelength-independent factor always makes the apparent reflectance spectrum different from the standard spectrum measured in a laboratory even if the materials are the same. Such a factor appears as an offset in the natural logarithm of reflectance

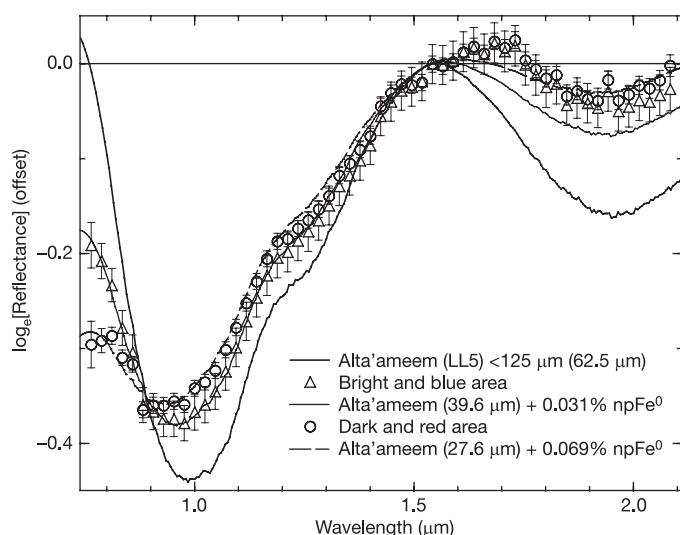


Figure 4 | Estimation of the degrees of space weathering of two areas on Itokawa. Natural log reflectance spectra of dark and red area and bright and blue area of Tsukuba are fitted with the spectrum of Alta'ameem (LL5) <125 μm sample by optimizing the mean optical path length (MOPL) and volume percentage of npFe^0 particles. Considering the data quality and the importance of the 1- μm band feature as indicators of the degree of space weathering, grain size, and composition, only the NIRS data in the wavelength range from 0.76 to 1.00 μm are used in the optimization process, with the data point at 1.54 μm always offset to 0. Error bars are 1σ .

spectrum. The above offsetting process cancels that factor. Also, there seems to be a clear calibration problem around 1.7 μm . However, the fact that the dark and red area is more space-weathered than the bright and blue area should still hold true.

Some may also suspect that a similar optical effect (darkening and reddening) could result from enrichment of metallic iron particles on the surface regolith. If Itokawa is made of LL5 or LL6 ordinary chondrite, metallic iron grains cannot be separated unless the regolith contains many fine particles less than 100 μm in size. However, close-up images of AMICA reveal that the entire surface of Itokawa is full of much larger pebbles and boulders. Therefore, it is highly unlikely that metallic iron or other opaque mineral grains have separated from the silicate grains and have become concentrated on the surface.

Because space weathering process is believed to take a significantly longer time than the average surface lifetime of a small body such as Itokawa, a surface that can accumulate the effect of space weathering needs to be stable and have very little new material falling over it. Sagami-hara may be one such place on Itokawa. The features on

asteroid Itokawa observed by the Hayabusa spacecraft may present the first clear look at an asteroid of ordinary-chondrite composition during regolith formation and while developing space weathering.

Received 21 April; accepted 6 July 2006.

- Adams, J. B. & McCord, T. B. Alteration of lunar optical properties: age and composition effects. *Science* **171**, 567–571 (1971).
- Binzel, R. P. *et al.* Observed spectral properties of near-Earth objects: Results for population distribution, source regions, and space weathering processes. *Icarus* **170**, 259–294 (2004).
- Abe, M. *et al.* Near-infrared spectral results of asteroid Itokawa from the Hayabusa spacecraft. *Science* **312**, 1334–1338 (2006).
- Hapke, B. Space weathering from Mercury to the asteroid belt. *J. Geophys. Res.* **106**, 10039–10073 (2001).
- Saito, J. *et al.* Detailed images of asteroid 25143 Itokawa from Hayabusa. *Science* **312**, 1341–1344 (2006).
- Ishiguro, M. *et al.* Detection of a large variation in the degree of space weathering on the surface of Itokawa by Hayabusa/AMICA observations. *Lunar Planet. Sci.* **37**, abstr. 1533 (2006).
- Binzel, R. P. *et al.* MUSES-C target asteroid (25143) 1998 SF36: A reddened ordinary chondrite. *Meteorit. Planet. Sci.* **36**, 1167–1172 (2001).
- McSween, H. Y., Bennett, M. E. III & Jarosewich, E. Mineralogy of ordinary chondrites and implications for asteroid spectrophotometry. *Icarus* **90**, 107–116 (1991).
- Klein, C. & Hurlbut, C. S. (eds) *Manual of Mineralogy*, 20th edn, 375 (Wiley, New York, 1985).
- Hiroi, T. & Pieters, C. M. Estimation of grain sizes and mixing ratios of fine powder mixtures of common geologic minerals. *J. Geophys. Res.* **99**, 10867–10879 (1994).
- Johnson, R. & Christy, R. Optical constants of transition metals: Ti, V, Cr, Mn, Fe, Co, Ni and Pd. *Phys. Rev. B* **9**, 5056–5070 (1974).
- Abe, S. *et al.* Mass and local topography measurements of Itokawa by Hayabusa. *Science* **312**, 1344–1347 (2006).
- Gaskell, R. *et al.* Global topography of asteroid 25143 Itokawa. *Lunar Planet. Sci.* **37**, abstr. 1876 (2006).
- Al-Bassam, K. S. The mineralogy and chemistry of the Alta'ameem meteorite. *Meteoritics* **13**, 257–265 (1978).

Acknowledgements The Alta'ameem meteorite sample (USNM 5964) was loaned from the National Museum of Natural History, Smithsonian Institution. We thank T. McCoy for assistance. The reflectance spectrum of the Alta'ameem sample was obtained at the Reflectance Experiment Laboratory (RELAB), a multiuser facility operated under the NASA Planetary Geology and Geophysics Program. The manuscript benefited from a review by B. Hapke.

Author Contributions T.H. contributed to the overall analyses and measurements of some meteorite samples in this work. M.A., K.K. and S.A. contributed to production and calibration of NIRS data as well as discussion on the NIRS data quality. K.K. and B.E.C. contributed to the photometric correction of NIRS spectra. B.E.C. and O.S.B.-J. contributed to deriving footprint positions of NIRS observations. S.A. provided gravity information. M.I. contributed to choosing the target location on Itokawa to study in this work based on AMICA colour image works. S.S. contributed to providing meteorite samples and space weathering discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.H. (takahiro_hiroi@brown.edu).

Finite lifetime of turbulence in shear flows

Björn Hof^{1,2}, Jerry Westerweel², Tobias M. Schneider³ & Bruno Eckhardt³

Generally, the motion of fluids is smooth and laminar at low speeds but becomes highly disordered and turbulent as the velocity increases. The transition from laminar to turbulent flow can involve a sequence of instabilities in which the system realizes progressively more complicated states¹, or it can occur suddenly^{2,3}. Once the transition has taken place, it is generally assumed that, under steady conditions, the turbulent state will persist indefinitely. The flow of a fluid down a straight pipe provides a ubiquitous example of a shear flow undergoing a sudden transition from laminar to turbulent motion^{4–6}. Extensive calculations^{7,8} and experimental studies⁹ have shown that, at relatively low flow rates, turbulence in pipes is transient, and is characterized by an exponential distribution of lifetimes. They^{8,9} also suggest that for Reynolds numbers exceeding a critical value the lifetime diverges (that is, becomes infinitely large), marking a change from transient to persistent turbulence. Here we present experimental data and numerical calculations covering more than two decades of lifetimes, showing that the lifetime does not in fact diverge but rather increases exponentially with the Reynolds number. This implies that turbulence in pipes is only a transient event (contrary to the commonly accepted view), and that the turbulent and laminar states remain dynamically connected, suggesting avenues for turbulence control¹⁰.

An abrupt transition to turbulence occurs in a large variety of situations, ranging from the formation of spiral galaxies, to the fluid motion in the atmosphere and oceans, to flows over aeroplane wings or through blood vessels. The flow down a circular pipe is a prime example, and an experimentally accessible realization, where such a transition occurs despite the linear stability of the base flow¹¹. We have recently developed a description for the onset of turbulence in pipes that focuses on the appearance of unstable travelling waves as the dimensionless flow rate—the Reynolds number, Re —is increased^{3,12–14} (here $Re = UD/\nu$, where U is the mean flow speed, D the pipe diameter and ν the kinematic viscosity of the fluid). These new solutions show up transiently in the turbulent signal, and are thought to connect to form a chaotic saddle^{15,16}. The most significant evidence for the formation of a chaotic saddle is the observation of the decay of turbulence at some point in time without any noticeable prior indication^{8,9,17}. If the probability of decay is constant in time and independent of when the turbulence was started, then the distribution of lifetimes is exponential, as in radioactive decay. Evidence for the presence of a chaotic saddle comes from the observation of such exponential distributions in pipe flow^{8,9} and in plane Couette^{18,19} flow (the flow of a fluid layer between two moving walls).

Formally, the probability of finding a turbulent flow state at a time t as a function of the Reynolds number can be expressed as:

$$P(t, Re) = \exp[-(t - t_0)/\tau(Re)] \quad (1)$$

where t is the time since the start of the experiment, t_0 a constant delay related to the initial formation of a turbulent flow state ($t_0 \ll \tau$), and τ is the characteristic lifetime, which depends on Re . The very rapid increase of $\tau(Re)$ with Re reported in earlier studies^{8,9,18} led to the conclusion that τ diverges at some finite Reynolds number Re_{crit} , that is, $\tau(Re) \propto 1/(Re_{crit} - Re)$. Such a divergence would imply that the system undergoes a transition from a transient chaotic saddle to a permanent chaotic attractor. Here we will reconsider this conclusion in view of new experimental and numerical results for observation times that span more than two orders of magnitude.

On the experimental side, a 30-m-long pipe was constructed from 20 precision bore copper segments, each 1.5 m long with an internal diameter of $D = 4$ mm (± 0.06 mm): this gives a maximal length to diameter ratio of 7,500. The flow was gravity driven, thus allowing an extremely accurate control over the flow rate (better than 0.1%), which is a crucial requirement for lifetime measurements. The friction factor of the laminar flow accurately follows the Hagen-Poiseuille law (see Supplementary Fig. S1, and Supplementary Methods), confirming the quality of the experimental set-up. Turbulent structures (so-called turbulent puffs^{20,21}) were created by injecting a water jet for 0.8 s through one of twenty 0.6 mm holes in the pipe wall; these holes were equally spaced along the pipe. The turbulent structure thereby created travels down the pipe at approximately the bulk flow speed^{20,22}, and the observation time is set by the distance between the chosen perturbation point and the pipe end. The pipe length of $7,500D$ allows a maximum observation time of $t = 7,500 D/U$: this is ten times longer than the maximal observation times in the currently longest laboratory pipe facilities, and ten times that reached in numerical calculations. As the small diameter of the pipe makes velocity measurements (such as particle image velocimetry or laser Doppler measurements) difficult, a much simpler but equally effective way of distinguishing laminar from turbulent flow²³ was chosen here. Because the centre-line velocity of turbulent flow is reduced by $\sim 30\%$ with respect to the laminar profile while the velocities close to the wall increase, turbulent and laminar flows leave the pipe at distinctly different angles (Fig. 1). By monitoring the jet leaving the pipe, we can therefore determine if a turbulent puff created upstream has survived the journey downstream.

In previous studies^{8,9,18}, the probability $P(t, Re)$ was measured at fixed Reynolds numbers as a function of time t . In the present set-up, it was more practical to keep the dimensionless observation time, L/D (that is, the distance between the point of perturbation and the outlet), fixed, and to vary the Reynolds number. From equation (1) and an assumed divergence of the characteristic lifetime, $\tau(Re) \propto 1/(Re_{crit} - Re)$, we would expect that $P(t, Re)$ increases exponentially with Re , that the steepness of the exponential increases with t and that independent of the value of t all curves reach $P = 1$ at Re_{crit} (Fig. 2a inset). Clearly, our measurements (Fig. 2a) do not support this

¹School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK. ²Laboratory for Aero & Hydrodynamics, Delft University of Technology, 2628 CA Delft, The Netherlands. ³Fachbereich Physik, Philipps-Universität Marburg, 35032 Marburg, Germany.

expectation. The data fall along S-shaped curves, which bend in the direction of larger Re as P increases towards 1. Moreover, by appropriate shifts in Re , the data for different observation times can be collapsed onto a single curve (Fig. 2b). The shape of this curve clearly does not show the steep rise expected for lifetimes diverging at a finite Re_{crit} .

The possibility of mapping the observations for different times onto one curve by shifting the Reynolds number suggests an exponential relation between the characteristic time and the Reynolds number, of the form $\tau^{-1} = \exp(a + bRe)$, with constants a and b . (Note that here and in the following all times are presented in their dimensionless form, unless stated otherwise.) This functional dependence implies that the shifts ΔRe needed to map the probability distributions for observation times t_1 and t_2 on top of each other are given by $\Delta Re = (1/b) \ln[(t_1 - t_0)/(t_2 - t_0)]$. This is indeed the case (Fig. 2b inset), and confirms the exponential scaling of τ . Alternatively, $\tau(Re)$ can be obtained directly from the data in Fig. 2a, and these verify the exponential scaling obtained above, as shown in Fig. 3a. From these fits, we obtain a good representation of our data with $b = -0.0344$, $a = 59.28$ and $t_0 = 120$. Substituting these parameters into equation (1) gives the theoretical curves shown in Fig. 2a, which reproduce the S-shape of $P(t, Re)$. In conclusion, the S-shape of the distributions in Fig. 2a, the scaling of shifts required to overlap these curves, the absence of divergence at a critical Reynolds number, as well as the characteristic lifetimes shown in Fig. 3a, are all fully consistent with the exponential scaling proposed for τ^{-1} . Clearly these features cannot be explained by the linear scaling proposed in earlier studies.

Direct numerical simulations of flow in pipes cannot at present reach these long observation times, and are limited by the extent of the numerical domain. Nevertheless, simulations that we performed for a length of $5D$ with periodic boundary conditions and for observation times up to $t = 3,000 R/U_{cl}$ (here R is the pipe radius and U_{cl} the centre-line velocity) agree with the experimental observations, and support the conclusion of transient turbulence with a

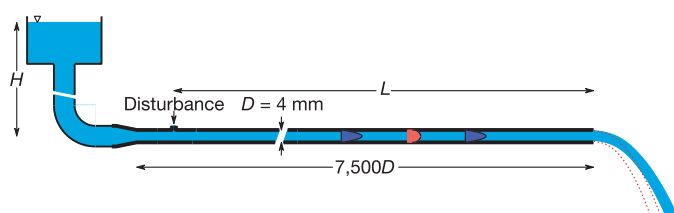


Figure 1 | Sketch of the experimental apparatus. The pipe sections were accurately aligned with a laser, ensuring that deviations were smaller than 2 mm over the entire 30 m length ($30 \text{ m} = 7,500D$, where D is the internal diameter, 4 mm). The flow is driven by a constant pressure head (H) and remained laminar up to Reynolds numbers greater than 2,750, where perturbations at the inlet become large enough to trigger turbulence; this is well above the largest Reynolds number of our measurements. During each measurement, a jet is injected into the laminar flow at a distance L from the pipe entrance with the aid of a high-speed solenoid valve. This perturbation creates a turbulent patch of fluid, a so-called turbulent puff. Gravity feeding ensures a constant pressure difference along the pipe, but the flow rate will drop when turbulence sets in, owing to the increased drag. However, all our studies are limited to a regime where the turbulent structures are smaller than $20D$ in length and their size remains constant^{17,21}. Theoretical estimates and experimental observations show that the maximum decrease in the flow rate caused by a single turbulent event is smaller than 0.25%. The turbulent puff travels downstream at the mean velocity towards the pipe exit where the turbulent flow (red velocity profile) can be distinguished from the laminar one (blue profiles) by a distinct drop in the outflow angle (shown dashed). The fluid was then pumped back into the header tank, keeping H constant to within 0.1%. The distance between the pipe entrance and the perturbation point could be varied in 20 steps of length $375D$, resulting in an equal change of the observation time, t .

characteristic lifetime that increases exponentially with Reynolds number (Fig. 3b). In these simulations, the flow was initially disturbed and integrated in time until the turbulent disturbance decayed or the maximum integration time of 3,000 was reached. For every Reynolds number, 100 independent simulation runs were analysed to extract the probability distribution of lifetimes (Supplementary Methods, Fig. S2), and τ^{-1} is again found to vary

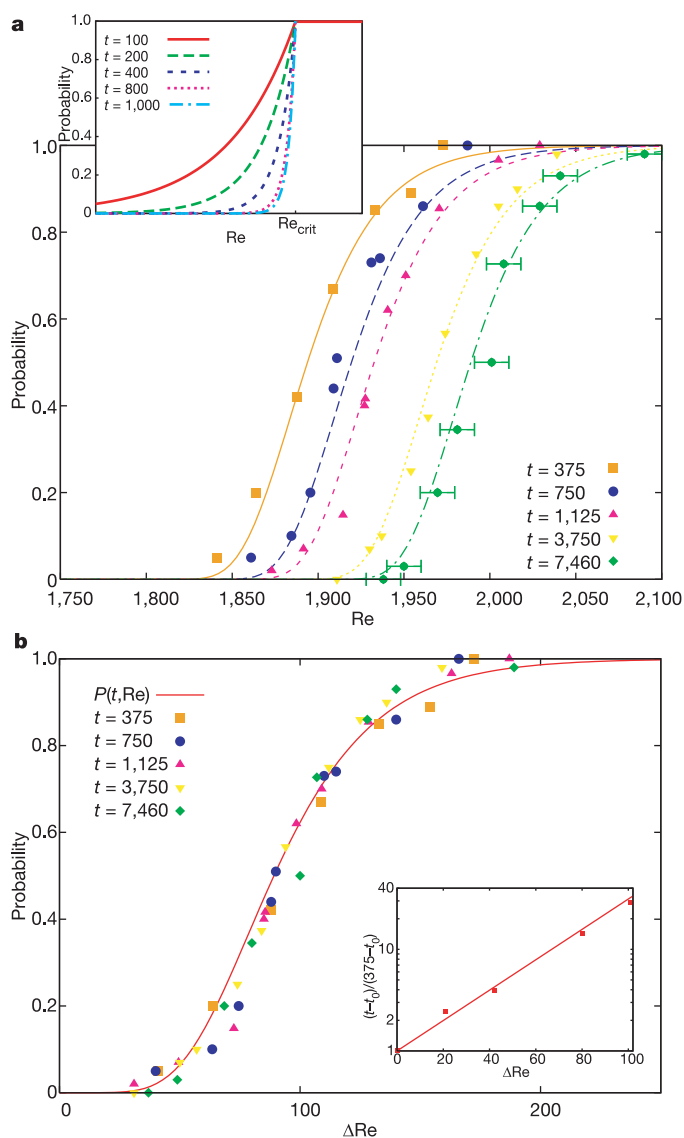


Figure 2 | Lifetime distributions. **a**, Probability distributions of lifetimes measured experimentally for fixed values of t (time since the start of the experiment) as a function of Reynolds number, Re . Each data point has been obtained from 50 to 75 rehearsals of the experiment. The distributions are S-shaped, flattening when approaching probability $P = 1$. With increasing t , the distribution curves shift to larger Re . The fitted lines are of the form $\exp[-(t - t_0)/\tau]$ with $\tau^{-1} = \exp(a + bRe)$; here t_0 is a constant delay, τ is the characteristic lifetime, and a and b are fitting parameters. The error bars represent the absolute uncertainty in the Reynolds number (for details see Supplementary Methods). Inset, theoretical predictions for a linear scaling of $1/\tau$ with Re . **b**, Same data as in **a** but with the curves collapsed by appropriate shifts ΔRe in Reynolds number, confirming that the shape of the distributions does not change notably with increasing t . The fitted line is obtained with the fitting parameters a and b from panel **a**. Inset, the ratio of observation times as a function of ΔRe . The exponential correlation is in accord with the suggested exponential variation of lifetimes with Re . The straight line is obtained with the fitting parameters a and b from panel **a**.

exponentially with Re (Fig. 3b). The main difference between experiments and the simulation is the finite length of the simulated pipe. Additional simulations of a longer pipe indicate that the same lifetimes are reached at lower Re .

Our observation of non-diverging lifetimes can be shown to be fully consistent with the data of previous studies, where a different conclusion had been drawn: reanalysing the numerical data of ref. 8

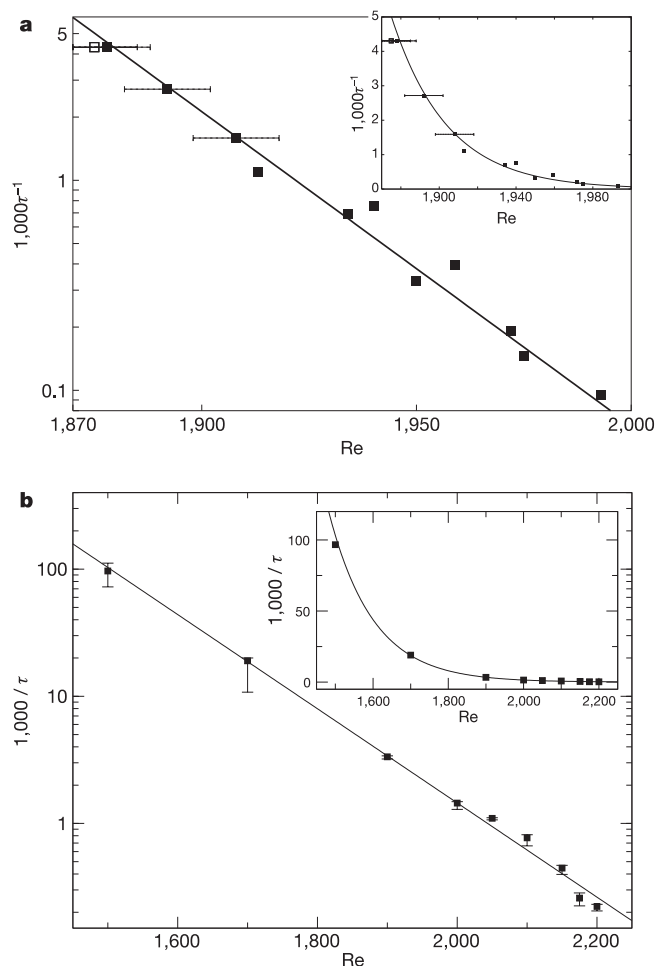


Figure 3 | Variation of characteristic times with Reynolds number.

a, The reciprocal characteristic lifetimes as a function of Re for all experiments with $t_0 = 120D/U$. They were obtained from fits to the data shown in Fig. 2a. Each data point required 400 to 500 measurements, with the total number of experiments underlying the figure exceeding 5,000. The straight line is an exponential fit to the data points. Error bars represent absolute errors based on experimental uncertainties (see Supplementary Methods for details) and are representative for all data points. Lifetimes were observed to be independent of the amplitude of the perturbation that initially created the turbulent patch. Typically, the volume flux injected at the perturbation point to trigger turbulence was 0.07 times the volume flux in the pipe. The open symbol presents a set of measurements where turbulent puffs were created by perturbation amplitudes four times larger than for the neighbouring symbol; both measurements agree within the experimental uncertainties. Inset, the same data replotted on a linear scale, underlining that they are not compatible with a diverging characteristic lifetime (error bars are the same as in the main figure). **b**, Characteristic lifetimes obtained from numerical simulations of pipe flow. The values of τ have been extracted from the slopes of the exponential tails of the probability density function, PDF, of lifetimes (see also material shown in Supplementary Fig. S2), and are plotted as a function of Re on a logarithmic scale and on a linear scale (inset). The curves correspond to an exponential fit to the data. The data clearly support an exponential scaling of τ as a function of the Re . The error bars represent the statistical uncertainties of the exponential fits. More precisely they show the range of values compatible with the calculated PDF.

by taking a non-zero t_0 into account reveals that lifetimes indeed vary exponentially with Reynolds number. Concerning the experimental lifetime measurements⁹, it is apparent from Fig. 3a inset that an exponential scaling of τ^{-1} can only be distinguished from a linear one if observation times exceed $t = 1,500$; this is much longer than the 500 available in ref. 9. Turning to other systems, evidence for a non-diverging lifetime can be found in calculations for Couette flow²⁴, where $P(t, Re)$ also shows the characteristic S-shaped form (see Supplementary Methods, Fig. S4). Similarly, the experimental data for plane Couette flow¹⁸ (see Supplementary Methods, Fig. S3) also show an exponential variation of the characteristic lifetimes once t_0 is taken into account. Finally, we mention that exponential behaviour is also found in a shear flow model²⁵.

We have shown that in contrast to previous findings, the lifetimes of turbulence in pipe flow do not diverge at a finite critical Reynolds number. All the available data indicate that turbulence in pipe flows decays, and that eventually the flow will always relaminarize. Similarly, lifetimes for Couette^{18,19} flow and for a shear flow model²⁵ do not appear to diverge, suggesting that a finite lifetime of turbulence may be a universal property of this class of flows. The rapid exponential increase of lifetimes explains why the transient nature of turbulence has not been observed previously: to detect the decay of turbulence in a garden hose at a flow rate as low as 1 l min^{-1} ($Re = 2,400$) would require a physical length of the tube of 40,000 km, about the Earth's circumference, and an observation time of almost 5 years. The dynamical connection between the turbulent and the laminar state implies that the flow can be relaminarized by infinitesimal perturbations in judiciously chosen directions. This implies in particular that turbulence control should be possible with minimal expenditures in energy.

Received 1 June; accepted 19 July 2006.

- Niemiela, J. J., Skrbek, L., Sreenivasan, K. R. & Donnelly, R. Turbulent convection at very high Rayleigh numbers. *Nature* **404**, 837–840 (2000).
- Grossmann, S. The onset of shear flow turbulence. *Rev. Mod. Phys.* **72**, 603–618 (2000).
- Hof, B. *et al.* Experimental observation of nonlinear travelling waves in turbulent pipe flow. *Science* **305**, 1594–1598 (2004).
- Reynolds, O. An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous and of the law of resistance in parallel channels. *Proc. R. Soc. Lond.* **35**, 84–99 (1883).
- Kerswell, R. R. Recent progress in understanding the transition to turbulence in a pipe. *Nonlinearity* **18**, R17–R44 (2005).
- Eckhardt, B., Schneider, T. M., Hof, B. & Westerweel, J. Turbulence transition in pipe flow. *Annu. Rev. Fluid Mech.* (submitted).
- Brosa, U. Turbulence without strange attractor. *J. Stat. Phys.* **55**, 1303–1312 (1989).
- Faisst, H. & Eckhardt, B. Sensitive dependence on initial conditions in transition to turbulence in pipe flow. *J. Fluid Mech.* **504**, 343–352 (2004).
- Peixinho, J. & Mullin, T. Decay of turbulence in pipe flow. *Phys. Rev. Lett.* **96**, 094501 (2006).
- Shinbrot, T., Grebogi, C., Ott, E. & Yorke, J. A. Using small perturbations to control chaos. *Nature* **363**, 411–417 (1993).
- Drazin, P. G. & Reid, W. H. *Hydrodynamic Stability* (Cambridge Univ. Press, Cambridge, UK, 1981).
- Faisst, H. & Eckhardt, B. Travelling waves in pipe flow. *Phys. Rev. Lett.* **91**, 224502 (2003).
- Eckhardt, B. & Mersmann, A. Transition to turbulence in a shear flow. *Phys. Rev. E* **60**, 509–517 (1999).
- Wedin, H. & Kerswell, R. R. Exact coherent solutions in pipe flow: travelling wave solutions. *J. Fluid Mech.* **508**, 333–371 (2004).
- Kadanoff, L. P. & Tang, C. Escape from strange repellers. *Proc. Natl Acad. Sci. USA* **81**, 1276–1279 (1984).
- Kantz, H. & Grassberger, P. Repellers, semi-attractors, and long-lived chaotic transients. *Physica D* **17**, 75–86 (1985).
- Hof, B. in *Proc. IUTAM Symp. on Laminar-Turbulent Transition and Finite Amplitude Solutions* (eds Mullin, T. & Kerswell, R. R.) 221–231 (Springer, Dordrecht, 2005).
- Bottin, S. & Chaté, H. Statistical analysis of the transition to turbulence in plane Couette flow. *Eur. Phys. J. B* **6**, 143–155 (1998).
- Schmiegel, A. & Eckhardt, B. Fractal stability border in plane Couette flow. *Phys. Rev. Lett.* **79**, 5250–5253 (1997).
- Wygnanski, I. J. & Champagne, F. H. On transition in a pipe. Part 1. The origin

- of puffs and slugs and the flow in a turbulent slug. *J. Fluid Mech.* **59**, 281–335 (1973).
21. Wygnanski, I. J., Sokolov, M. & Friedman, D. On transition in a pipe. Part 5. The equilibrium puff. *J. Fluid Mech.* **69**, 283–304 (1975).
 22. Hof, B., vanDoorne, C. W. H., Westerweel, J. & Nieuwstadt, F. T. M. Turbulence regeneration at moderate Reynolds numbers. *Phys. Rev. Lett.* **95**, 214502 (2005).
 23. Rotta, J. C. Experimenteller Beitrag zur Entstehung turbulenter Strömungen im Rohr. *Ing. Archiv* **24**, 258–281 (1956).
 24. Schmieg, A. *Transition to Turbulence in Linearly Stable Shear Flows*. PhD thesis, Philipps-Univ. Marburg (1999).
 25. Moehlis, J., Faisst, H. & Eckhardt, B. A low-dimensional model for turbulent shear flows. *N. J. Phys.* **6**, 56 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded through a RCUK fellowship (B.H.), the Foundation of Fundamental Research on Matter (J.W.) and Deutsche Forschungsgemeinschaft (T.M.S. and B.E.). B.H. and J.W. thank R. Delfos for discussions and W. Tax and P. Tipler for technical assistance. We thank the Dorset Tube Company for supplying the precision pipe.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.H. (bjorn@reynolds.ph.man.ac.uk) or J.W. (j.westerweel@tudelft.nl).

A class of non-precious metal composite catalysts for fuel cells

Rajesh Bashyam¹ & Piotr Zelenay¹

Fuel cells, as devices for direct conversion of the chemical energy of a fuel into electricity by electrochemical reactions, are among the key enabling technologies for the transition to a hydrogen-based economy^{1–3}. Of several different types of fuel cells under development today, polymer electrolyte fuel cells (PEFCs) have been recognized as a potential future power source for zero-emission vehicles^{4,5}. However, to become commercially viable, PEFCs have to overcome the barrier of high catalyst cost caused by the exclusive use of platinum and platinum-based catalysts^{6–8} in the fuel-cell electrodes. Here we demonstrate a new class of low-cost (non-precious metal)/(heteroatomic polymer) nanocomposite catalysts for the PEFC cathode, capable of combining high oxygen-reduction activity with good performance durability. Without any optimization, the cobalt-polypyrrole composite catalyst enables power densities of about 0.15 W cm^{-2} in $\text{H}_2\text{-O}_2$ fuel cells and displays no signs of performance degradation for more than 100 hours. The results of this study show that heteroatomic polymers can be used not only to stabilize the non-precious metal in the acidic environment of the PEFC cathode but also to generate active sites for oxygen reduction reaction.

A schematic representation of the principle of fuel-cell operation is shown in Fig. 1. PEFCs operate with a polymer electrolyte membrane that separates the fuel (hydrogen) from the oxidant (air or oxygen). Precious-metal catalysts, predominantly platinum (Pt) supported on carbon, are used for both the oxidation of the fuel and reduction of the oxygen in a typical temperature range of $80\text{--}100^\circ\text{C}$. Apart from the issue of the high cost of catalyst and other fuel-cell system components (polymer electrolyte membrane, bipolar plates, the rest of the power system, and so on), PEFCs suffer from insufficient performance durability, arising mainly from cathode catalyst oxidation, catalyst migration, loss of electrode active surface area, and corrosion of the carbon support. In a direct methanol fuel cell (DMFC), the Pt cathode also endures a performance loss resulting from the lack of tolerance to methanol diffusing through the membrane from the anode side of the cell^{9,10}. Thus, whether using hydrogen or methanol as a fuel, PEFCs are in need of efficient, durable and, most importantly, inexpensive catalysts, as alternatives to Pt and Pt-based materials^{11,12}. Although ideally the Pt catalyst should be replaced at both fuel-cell electrodes, the substitution of the cathode catalyst with a non-precious material is likely to result in significantly greater reduction of Pt needed for PEFCs. This is because the slow oxygen reduction reaction (ORR) at the cathode requires much more Pt catalyst than the very fast hydrogen oxidation at the anode¹³.

Two approaches are at present gaining momentum to replace Pt, which is scarce (only 37 p.p.b. in the Earth's crust) and expensive ($\sim \$45 \text{ g}^{-1}$, the highest Pt price in 25 years). One approach uses non-Pt catalysts that nonetheless contain precious metals with limited abundance and/or limited world distribution. Such catalysts

are typically based on palladium (Pd; ref. 14) or ruthenium (Ru; refs 15, 16). Although Pt is thus avoided, the result is the replacement of one precious metal with another that is on the whole less active than Pt. An alternative approach is to replace Pt with abundant, non-precious materials that are not susceptible to price inflation under high-demand circumstances.

The class of non-precious cathode catalysts that has attracted the most attention over the years is pyrolyzed metal porphyrins, with cobalt and iron porphyrins viewed as the most promising precursors. The pyrolysis of various such porphyrins, carried out under controlled conditions of temperature, pyrolysis time and metal concentration in the porphyrins, has been shown to produce species of an MeN_x type ($\text{Me} = \text{Co}, \text{Fe}$), which have been suggested to be the active ORR sites^{17–19}. The hypothesis involving Me-N active sites has been supported by the studies correlating fuel-cell performance with nitrogen concentration in the porphyrin catalyst²⁰. Most probably owing to the high content of nitrogen, non-pyrolyzed metal porphyrins have been found to exhibit good initial oxygen reduction performance but dismal stability²¹.

The three-decade-long search for non-precious-metal catalysts for the PEFC cathode has so far revealed very few materials with

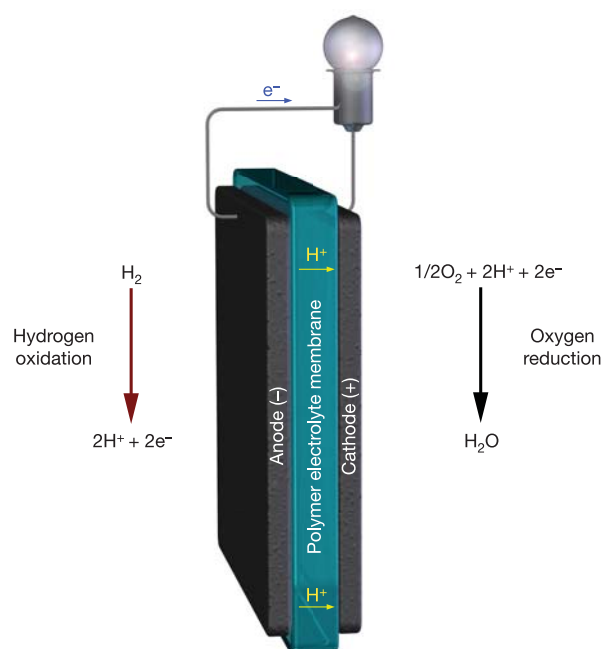


Figure 1 | Polymer electrolyte $\text{H}_2\text{-O}_2$ fuel cell.

¹Los Alamos National Laboratory, Materials Physics and Applications, Los Alamos, New Mexico 87545, USA.

promising ORR activity in either fundamental^{122–24} or in fuel-cell studies^{25,26}, and virtually none with sufficient performance stability. Even the most thoroughly studied pyrolyzed metal-porphyrins continue to lack durability under practical fuel-cell operating conditions, which re-emphasizes the need for more research on alternative non-precious cathode electrocatalysts for polymer electrolyte fuel cells²⁷.

Here we report major progress towards the development of a new class of inexpensive (non-precious metal)/(heteroatomic polymer) electrocatalysts for oxygen reduction in fuel cells. In particular, we demonstrate a cobalt-polypyrrole-carbon (Co-PPY-C) composite synthesized via a simple chemical method, without resorting to pyrolysis. The obtained catalyst shows good activity with a Co loading of $6.0 \times 10^{-2} \text{ mg cm}^{-2}$ and remarkable stability in both $\text{H}_2\text{-O}_2$ and $\text{H}_2\text{-air}$ PEFCs.

Recently, heterocyclic conjugated polymers such as polyaniline, polypyrrole and polythiophene have been the subject of much research owing to their applications in electronics, electrochemistry and electrocatalysis, as well as in biosensors and actuators^{28,29}. These applications primarily use the electronic conductivity of conjugated polymers. In this work, however, polypyrrole is solely used as a matrix for entrapping cobalt and generating active ORR sites. The

main reason for introducing polypyrrole into the structure of a composite ORR catalyst was to mimic the atomic configuration occurring in cobalt porphyrins, wherein the Co atom is linked to pyrrole units, thus allowing for the formation of Co-N sites. Consequently, the underlying assumption of this work was to purposely create active Co-N sites without destroying the initial polymer structure. This approach represents a departure from the pyrolysis process, which has been universally used until now to induce ORR activity in the synthesis of non-precious-metal ORR catalysts.

A schematic representation of the polymer structure and presumed configuration of the composite are shown in Fig. 2a and b, respectively. While Fig. 2a depicts the structure of polypyrrole used for entrapping cobalt atoms, the purpose of Fig. 2b is to highlight the importance of a linkage between polypyrrole and Co atoms, established via nitrogen atoms in the pyrrole units. To ensure electronic conductivity in the catalyst phase, the composite shown in Fig. 2b was synthesized on a support made of Vulcan XC 72 carbon (see 'Preparation of polypyrrole-loaded carbon and synthesis of the cobalt-polypyrrole-carbon composite' in the Supplementary Information for details).

The composite was characterized by surface area measurements (Brunauer–Emmett–Teller isotherm method) and thermogravimetric analysis (TGA). The Co loading in the composite was

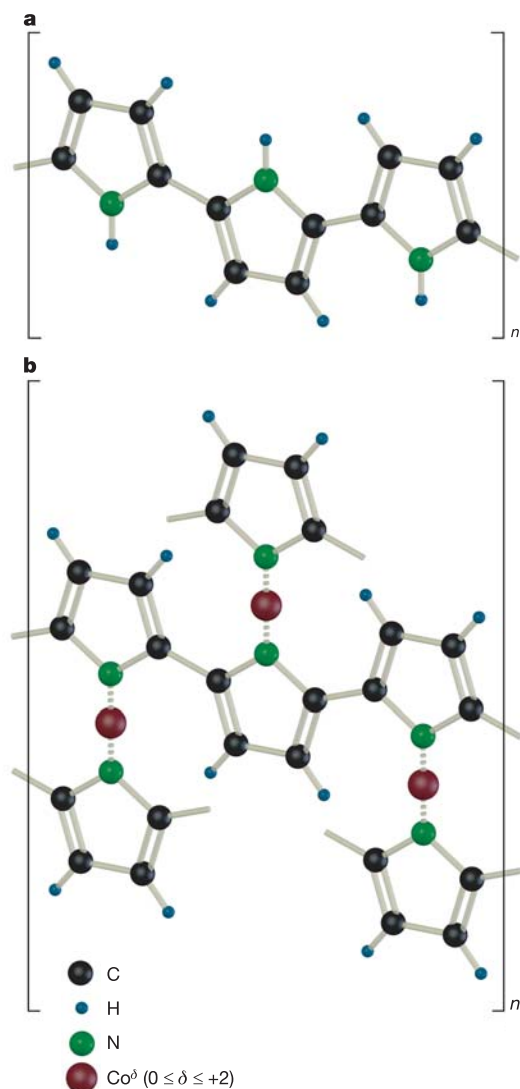


Figure 2 | Schematic representation of the Co-polypyrrole composite catalyst. a, Polypyrrole structure. **b**, Presumed configuration of the Co-PPY catalyst after the entrapment of the cobalt precursor, $\text{Co}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ into polypyrrole and reduction with NaBH_4 .

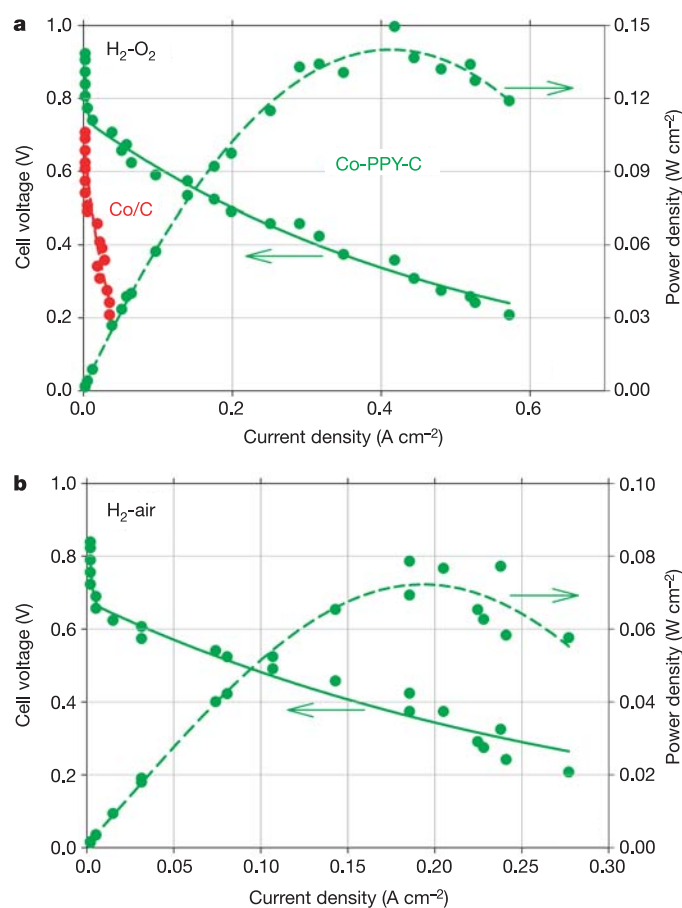


Figure 3 | Fuel-cell performance plots. a, Polarization and power density plots for $\text{H}_2\text{-O}_2$ fuel cells with Co-PPY-C composite cathode (green lines) and polarization plot for the cell with Co/C cathode (red line). The Co loading is $6.0 \times 10^{-2} \text{ mg cm}^{-2}$; the cell temperature is 80°C . **b**, Polarization and power density plots for $\text{H}_2\text{-air}$ fuel cells with Co-PPY-C composite cathode. The Co loading is $6.0 \times 10^{-2} \text{ mg cm}^{-2}$; the cell temperature is 80°C . Flow rates of hydrogen and oxygen/air at 5 ml s^{-1} and 9 ml s^{-1} (as referred to the standard conditions), respectively. Anode and cathode gases humidified at 90°C and 80°C , respectively. The backpressure of gases is 2.0 atm on both sides of the cell.

determined from TGA. At $124 \text{ m}^2 \text{ g}^{-1}$, the Brunauer–Emmett–Teller surface area of the Co-PPY-C composite was nearly 50% lower than the surface area of the Vulcan XC 72 support ($240 \text{ m}^2 \text{ g}^{-1}$), indicating that Vulcan XC 72 properties are modified by the polypyrrole. The complete decomposition of Vulcan XC 72 in the presence of oxygen occurs at 460°C , whereas that of Co-PPY-C composite occurs at 300°C . The TGA data suggest that oxidative decomposition of the composite is activated in the presence of cobalt. High-temperature decomposition notwithstanding, the composite catalyst remains stable in the temperature range of PEFC operation: between 80°C and 100°C . Based on the TGA, the polypyrrole content in the composite has been found to be nearly 16% by weight.

Membrane electrode assemblies for fuel-cell conditioning and testing of the new catalyst were fabricated with a Co-PPY-C cathode and Pt-Ru black anode (see experimental details in Supplementary Information). The membrane electrode assemblies were subject to conditioning at 0.40 V and at 80°C under $\text{H}_2\text{-O}_2$ fuel-cell operating conditions. The cell performance gradually increased during the conditioning step, reaching a constant level after approximately 30 h (see Supplementary Fig. 1). X-ray absorption near-edge structure (XANES) spectroscopy reveals that during conditioning most of the metallic Co transforms into Co(II) (S. Conradson, R.B. & P.Z., manuscript in preparation). X-ray absorption fine structure (XAFS) data indicate that stability of the Co(II) sites in the composite is achieved thanks to the formation of a coordination bond between Co(II) and N or O (the two elements are not easily distinguishable in XANES/XAFS).

An $\text{H}_2\text{-O}_2$ fuel-cell polarization plot recorded after the conditioning step with the cathode made of a newly synthesized Co-PPY-C composite is shown in Fig. 3a (see Supplementary Figs 2 and 3 for catalyst performance reproducibility and standard deviations, respectively). The catalyst generates $\sim 0.2 \text{ A cm}^{-2}$ at 0.50 V and a maximum power density of $\sim 0.14 \text{ W cm}^{-2}$. Although this current density is less than that recently reported in ref. 19 for an iron-based catalyst (0.3 A cm^{-2} at 0.50 V), it was obtained at a significantly lower absolute pressure of oxygen (2.8 atm versus the 5.0 atm used in ref. 19).

The $\text{H}_2\text{-O}_2$ fuel-cell performance of the composite catalyst compares favourably with other non-Pt cathode catalysts, even precious-metal-based materials, such as the recently proposed Pd-Co-Au alloy ($\sim 0.18 \text{ A cm}^{-2}$ at 0.50 V , 60°C , 1.5 atm oxygen back pressure) and Pd-Ti alloy (0.1 A cm^{-2} at 0.50 V , 60°C and 0.2 mg cm^{-2})¹⁴. Furthermore, the loading-independent oxygen reduction turnover rate of 0.83 s^{-1} at 0.50 V and 60°C (see 'Turnover rates' in the

Supplementary Information for details), is 3.5 times higher on the Co-PPY-C composite than on the Pd-Co-Au alloy (0.24 s^{-1}) and 8.3 times higher than on the Pd-Ti alloy (0.10 s^{-1}). The comparison becomes even more favourable for the composite when it is made on the basis of combining the ORR activity with performance durability (see below).

For a reference, a catalyst made by just loading Co onto Vulcan XC 72 was synthesized under the same conditions as the composite catalyst. It is evident from Fig. 3a that activity of the Co/C catalyst is far lower than that of the Co-PPY-C composite. The open cell voltage (that is, the voltage at zero cell current, OCV) of the fuel cell with the Co-PPY-C cathode is 0.22 V higher than that of the fuel cell with the Co/C cathode. In the voltage range between 0.50 V and 0.20 V , current density generated by the Co-PPY-C cell is more than one order of magnitude higher than that delivered by the Co/C cell (Fig. 3a). In spite of having a $\sim 50\%$ catalytic surface area advantage over the Co-PPY-C composite, the peak power density of the Co/C cell is only 0.009 W cm^{-2} , nearly 16 times lower than that of the Co-PPY-C cell.

The above data provide indirect evidence that electrocatalytic activity of the new catalyst results from the entrapment of the Co sites in the polypyrrole matrix and strong Co–polypyrrole interactions, presumably resulting in the formation of CoN_x active sites. This interaction between Co and polypyrrole is analogous to the interaction occurring between heterocyclic compounds (in particular heterocyclic polymers) placed in contact with metals. Such interaction is known to lead to a charge transfer to the metal, and, consequently, to the formation of a strong heterocycle–metal bonding³⁰.

An $\text{H}_2\text{-air}$ fuel-cell polarization plot recorded with a Co-PPY-C composite cathode is shown in Fig. 3b. In spite of the reduced partial pressure of O_2 in air, the cell performance is good, as reflected by the current density of 0.10 A cm^{-2} at 0.50 V . The maximum peak power density of the $\text{H}_2\text{-air}$ cell is $\sim 0.07 \text{ W cm}^{-2}$, approximately half of that generated with the $\text{H}_2\text{-O}_2$ cell (see above).

The most striking property of the new catalyst—or, more likely, of a new class of catalysts represented by the Co-PPY-C catalyst—is demonstrated in Fig. 4. Uncharacteristically for a non-precious catalyst, cell performance with the Co-PPY-C composite cathode is very stable, showing no appreciable drop over 100 h of operation. To the best of our knowledge, this represents the first time that a non-precious cathode catalyst, at a low loading, has shown both a promising activity and stability comparable to that of Pt-based ORR catalysts (compare with the 100-h life test of an E-TEK 20 wt% Pt/C catalyst; Supplementary Fig. 4). The good performance stability of the composite can be seen not only at the relatively low voltage of 0.40 V but also at more practical higher voltages, up to 0.70 V (see Supplementary Figs 5 and 6).

Fundamental electrochemical studies are underway to understand the mechanism of O_2 reduction on the composite materials, such as the Co-PPY-C composite catalyst introduced here. The main goal of future research will be to reduce the overpotential of oxygen reduction at high cathode operating potentials (in the low-current-density region) by at least $0.15\text{--}0.20 \text{ V}$ and to bring it closer to that measured with pure-Pt and Pt-based cathode catalysts.

To conclude, here we have demonstrated a new non-precious composite catalyst from an entirely new class of (non-precious metal)/(heterocyclic polymer) composite materials, synthesized via a pyrolysis-free process. As shown with the Co-PPY-C composite, such catalysts promise to deliver high ORR activity without any noticeable loss in performance over long fuel-cell operating times. This therefore opens up the possibility of making a variety of other non-precious composite materials from this class of composites for use as catalysts for the PEFC cathode.

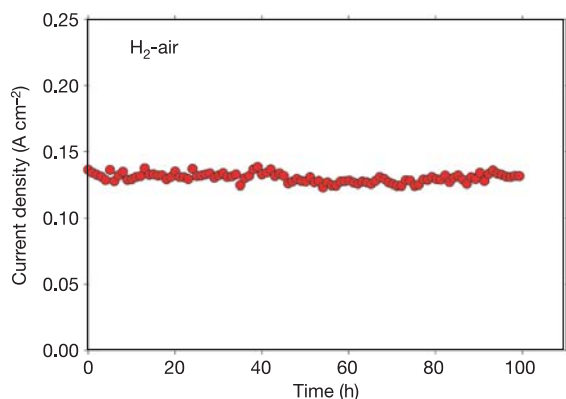


Figure 4 | Durability test of the composite catalyst. Long-term performance of an $\text{H}_2\text{-air}$ fuel cell with Co-PPY-C composite cathode at 0.40 V . The Co loading is $6.0 \times 10^{-2} \text{ mg cm}^{-2}$; the cell temperature is 80°C . Flow rates of hydrogen and air at 5 ml s^{-1} and 9 ml s^{-1} (as referred to the standard conditions), respectively. Anode and cathode gases humidified at 90°C and 80°C , respectively. The backpressure of gases is 2.0 atm on both sides of the cell.

Received 7 March; accepted 28 July 2006.

- Jacobson, M. Z., Colella, W. G. & Golden, D. M. Atmospheric science: Cleaning the air and improving health with hydrogen fuel-cell vehicles. *Science* **308**, 1901–1905 (2005).
- Dusastre, V. Materials for clean energy. *Nature* **414**, 331 (2001).
- Schultz, M. G., Diehl, T., Brasseur, G. P. & Zittel, W. Air pollution and climate-forcing impacts of a global hydrogen economy. *Science* **302**, 624–627 (2003).
- Steele, B. C. H. & Heinzel, A. Materials for fuel-cell technologies. *Nature* **414**, 345–352 (2001).
- Brumfiel, G. Hydrogen cars fuel debate on basic research. *Nature* **422**, 104 (2003).
- Berger, D. J. Fuel cells and precious-metal catalysts. *Science* **286**, 49 (1999).
- Semelsberger, T. A. & Borup, R. L. Fuel effects on start-up energy and efficiency for automotive PEM fuel cell systems. *Int. J. Hydrogen Energy* **30**, 425–435 (2005).
- Xie, J., Wood, D. L., More, K. L., Atanassov, P. & Borup, R. L. Microstructural changes of membrane electrode assemblies during PEFC durability testing at high humidity conditions. *J. Electrochem. Soc.* **152**, A1011–A1020 (2005).
- Zhang, X., Filho, L. P., Torras, C. & Garcia-Valls, R. Experimental and computational study of proton and methanol permeabilities through composite membranes. *J. Power Sources* **145**, 223–230 (2005).
- Thomas, S. C., Ren, X. M., Gottesfeld, S. & Zelenay, P. Direct methanol fuel cells: Progress in cell performance and cathode research. *Electrochim. Acta* **47**, 3741–3748 (2002).
- Zhang, J. *et al.* Platinum monolayer on nonnoble metal-noble metal core-shell nanoparticle electrocatalysts for O₂ reduction. *J. Phys. Chem. B* **109**, 22701–22704 (2005).
- Zhang, J. *et al.* Platinum monolayer electrocatalysts for O₂ reduction: Pt monolayer on Pd(111) and on carbon-supported Pd nanoparticles. *J. Phys. Chem. B* **108**, 10955–10964 (2004).
- Appleby, A. J. Electrocatalysis of aqueous dioxygen reduction. *J. Electroanal. Chem.* **357**, 117–179 (1993).
- Fernandez, J. L., Raghuvier, V., Manthiram, A. & Bard, A. J. Pd-Ti and Pd-Co-Au electrocatalysts as a replacement for platinum for oxygen reduction in proton exchange membrane fuel cells. *J. Am. Chem. Soc.* **127**, 13100–13101 (2005).
- Alonso-Vante, N. A. & Tributsch, H. Energy conversion catalysis using semiconducting transition-metal cluster compounds. *Nature* **323**, 431–432 (1986).
- Alonso-Vante, N., Malakhov, I. V., Nikitenko, S. G., Savinova, E. R. & Kochubey, D. I. The structure analysis of the active centers of Ru-containing electrocatalysts for the oxygen reduction. An in situ EXAFS study. *Electrochim. Acta* **47**, 3807–3814 (2002).
- Wang, B. Recent development of non-platinum catalysts for oxygen reduction reaction. *J. Power Sources* **152**, 1–15 (2005).
- Villers, D., Jacques-Bedard, X. & Dodelet, J. P. Fe-based catalysts for oxygen reduction in PEM fuel cells. Pretreatment of the carbon support. *J. Electrochem. Soc.* **151**, A1507–A1515 (2004).
- Medard, C., Lefevre, M., Dodelet, J. P., Jaouen, F. & Lindbergh, G. Oxygen reduction by Fe-based catalysts in PEM fuel cell conditions: Activity and selectivity of the catalysts obtained with two Fe precursors and various carbon supports. *Electrochim. Acta* **51**, 3202–3213 (2006).
- Faubert, G. *et al.* Heat-treated iron and cobalt tetraphenylporphyrins adsorbed on carbon black: physical characterization and catalytic properties of these materials for the reduction of oxygen in polymer electrolyte fuel cells. *Electrochim. Acta* **41**, 1689–1701 (1996).
- Beck, F. Redox mechanism of chelate-catalyzed oxygen cathode. *J. Appl. Electrochem.* **7**, 239–245 (1977).
- Jasinski, R. New fuel cell cathode catalyst. *Nature* **201**, 1212 (1964).
- Atanasoski, R. *Novel Approach to Non-Precious Metal Catalysts*. Report VIIC9 (US DOE, 2005); (http://www.hydrogen.energy.gov/pdfs/progress05/vii_c_9_atanasoski.pdf).
- Yuasa, M. *et al.* Modifying carbon particles with polypyrrole for adsorption of cobalt ions as electrocatalytic site for oxygen reduction. *Chem. Mater.* **17**, 4278–4281 (2005).
- Lalande, G. *et al.* Physical, chemical and electrochemical characterization of heat-treated tetracarboxylic cobalt phthalocyanine adsorbed on carbon black as electrocatalyst for oxygen reduction in polymer electrolyte fuel cells. *Electrochim. Acta* **40**, 2635–2646 (1995).
- Lefevre, M. & Dodelet, J. P. Fe-based catalysts for the reduction of oxygen in polymer electrolyte membrane fuel cell conditions: Determination of the amount of peroxide released during electroreduction and its influence on the stability of the catalysts. *Electrochim. Acta* **48**, 2749–2760 (2003).
- Zelenay, P. *Non-Precious Metal Catalysts*. Report VIIC7 (US DOE, 2005); (http://www.hydrogen.energy.gov/pdfs/progress05/vii_c_7_zelenay.pdf).
- Lu, W. *et al.* Use of ionic liquids for pi-conjugated polymer electrochemical devices. *Science* **297**, 983–987 (2002).
- Hepel, M., Chen, Y. M. & Stephenson, R. Effect of the composition of polypyrrole substrate on the electrode position of copper and nickel. *J. Electrochem. Soc.* **143**, 498–505 (1996).
- Skotheim, T. A. *et al.* Highly ordered thin films of polyheterocycles: a synchrotron radiation study of polypyrrole and polythiophene Langmuir-Blodgett films. *Synth. Met.* **28**, C229–C236 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the US Department of Energy (Hydrogen, Fuel Cells and Infrastructure Technologies Program) and the Los Alamos National Laboratory for funding this work. We are grateful to E. Brosa, A. Burrell, J. Chlistunoff, F. Garzon, C. Johnston, R. Mukundan and J. Ramsey from the Los Alamos National Laboratory for discussions and technical help.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.Z. (zelenay@lanl.gov).

Enantioselective silyl protection of alcohols catalysed by an amino-acid-based small molecule

Yu Zhao¹, Jason Rodrigo¹, Amir H. Hoveyda¹ & Marc L. Snapper¹

Reliable, selective and environmentally friendly chemical transformations are crucial to the development of new therapeutics and the design of novel materials. Chiral catalysts that can be easily prepared and used to obtain organic molecules of high enantiomeric purity are critical to modern chemical synthesis¹. The development of protecting groups that shield reactive functionalities has also proved indispensable in the preparation of complex biologically active molecules². Here we present a chiral catalyst that promotes the enantioselective protection of a secondary alcohol as one of the most commonly used protected forms of an alcohol: a silyl ether. The catalyst is a small, simple molecule that can be prepared in three steps from commercial materials without the need for rigorously controlled conditions. Enantioselective silylations are performed with commercial silyl chlorides and produce yields of up to 96 per cent at an enantiomeric ratio of up to 98:2. Chiral catalysts for selective formation of commonly used protecting groups such as silyl ethers should significantly enhance the ability of chemical synthesis to deliver, in a more practical and efficient manner, important organic molecules.

The significance of handedness in nature renders the availability of chiral products in high enantiomeric purity a crucial objective of modern chemistry. In this context, the discovery of efficient chiral catalysts that can be used to prepare enantiomerically enriched compounds, are easily prepared, stable in air, can be recycled and afford the desired products with minimal toxic impurities has been actively pursued³. Another area of investigation has focused on introducing practical strategies for site-specific modifications of organic molecules. One approach involves moieties known as 'protecting groups', used to mask selected sites within a molecule; reagents that allow the installation of protecting groups have been under continuous development² and are frequently used in preparations of biologically active molecules⁴. Although an ideal synthesis should not require

masking of a functional group, protecting groups are, and will probably remain, critical to organic synthesis². Among the most widely used protecting groups for alcohols (a commonly occurring functional group found in organic molecules) are silyl ethers⁵.

The development of chiral catalysts and Si-based protecting groups has been largely pursued separately, but there are exceptions. One report presents a reagent that can resolve (that is, maximum yield is 50%) a limited number of secondary alcohols by a procedure that requires six days and generates silyl ethers in moderate enantiomeric purity: not higher than 85:15 enantiomeric ratio (e.r.) or 70% enantiomeric excess (e.e.)⁶. A recent disclosure describes a method for Cu-catalysed kinetic resolution of 2-pyridyl-substituted benzylic secondary alcohols (in up to 90:10 e.r.); the protocol requires stoichiometric amounts of specially designed enantiomerically pure trialkylsilyl hydrides (stereogenic at Si)⁷. Although chiral alcohol acylation^{8,9}, phosphorylation¹⁰ and sulphonylation¹¹ catalysts are known, and such units can in principle be viewed as protecting groups, silyl ethers are used far more frequently in synthesis². Another distinction of catalytic silylation is the absence of a biomimetic mechanistic blueprint.

A chiral catalyst for alcohol silylation can greatly increase the efficiency with which optically enriched organic molecules are prepared; the example in Fig. 1 illustrates this point. Silyl ether **4** is a valuable building block used to prepare several biologically active entities in the optically active form (for example, neocarzinostats, and so on

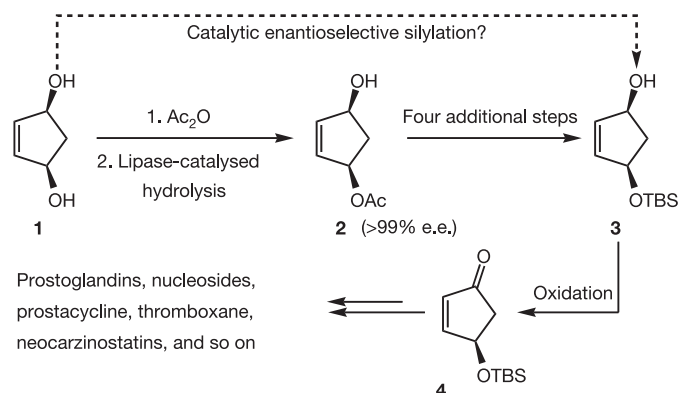


Figure 1 | A catalyst for enantioselective silylation of alcohols can render synthesis of cyclopentenone **4** significantly more efficient.

Table 1 | Initial screening of catalyst candidates for enantioselective silylation of substrate **1** (to produce **3**)

Entry number	Catalyst	Equivalent	e.r.	e.e. (%)
1	5	0.2	50:50	<5
2	6	0.2	58:42	16
3	7	0.2	50:50	<5
4	8	0.2	50:50	<5

The reactions were carried out in THF in the presence of 1.25 equivalents of DIPEA at 22 °C for 20–24 h and with a substrate concentration of 0.2 M.

In Tables 1 to 3 the enantiomeric ratio (e.r.) was determined by chiral gas liquid chromatography (GLC) analysis. See the Supplementary information for details. In Tables 1 to 3 the enantiomeric excess (e.e.) was calculated from the e.r.; the variance of e.e. are estimated to be $\pm 2\%$.

¹Department of Chemistry, Merkert Chemistry Center, Boston College, 2609 Beacon Street, Chestnut Hill, Massachusetts 02467, USA.

prostaglandins, thromboxane and nucleosides). Several procedures are known for the preparation of optically enriched **4**. One of the most efficient routes, in terms of overall yield, amount of waste produced and the time required, is the seven-step procedure shown in Fig. 1 (via **2** and **3**)¹². This sequence was developed to avoid another protocol that involves three chemical transformations for conversion of **2** to **4**, but demands an enzymatic (germ lipase) deacylation that consumes ten days: seven days for the enantioselective reaction, followed by a three-day procedure for isolation of the desired product that generates copious amounts of solvent waste. A chiral catalyst that directly converts achiral diol **1** to optically enriched silyl ether **3** by a catalytic enantioselective silylation would present a two-step sequence to optically enriched **4**. Thus, by shortening the synthesis route, waste production is minimized and time efficiency is enhanced. Although costly, the acetate to silyl ether 'swap' in Fig. 1, carried out through a four-step sequence, is significantly more efficient. This is because a silyl ether can be

removed selectively under milder conditions than an acetate or other ester groups, and without the risk of undesired side reactions that may occur under the conditions required for the removal of such groups. In contrast, a chiral acylation catalyst would reduce the sequence in Fig. 1 by only one step.

We considered two design criteria. One requirement was that the catalyst should contain a Lewis basic moiety that, through association, can increase the electrophilicity of silyl halide reagent (Lewis base activation^{13,14}). The other criterion, used previously in catalyst development^{15,16}, involves the use of H-bonding to establish substrate–catalyst binding¹⁷. We evaluated several representative molecules that were potential catalysts. The molecules selected carry potential H-bonding sites (for example, amide bonds) and nucleophilic groups that can interact with silyl chlorides (for example, *N*-oxide **5** (ref. 18), *N*-alkylimidazoles **6** and **7** (ref. 8) and pyridine **8** (ref. 19)).

This study was carried out in the context of a representative process: conversion of **1** to **3** with commercially available TBSCl

Table 2 | Catalytic enantioselective silylations of achiral diols

Entry number	Product	Catalyst	Catalyst equivalent	Time (h)	Temperature (°C)	e.r.	e.e. (%)	Yield (%)
1		9	0.2	120	−78	90.5:9.5	81	54
2		10	0.2	48	−78	93.5:6.5	87	55
3		9	0.3	48	−78	98:2	96	82
4		9	0.3	60	−40	94:6	88	96
5		9	0.2	120	−28	96:4	92	82
6		9	0.3	72	−40	97:3	94	75
7		9	0.3	72	−40	96.5:3.5	93	93
8		9	0.3	120	−40	97.5:2.5	95	96
9		9	0.3	72	−40	96.5:3.5	93	80
10		9	0.2	120	−28	95:5	90	84
11		9	0.3	72	−28	96:4	92	67

The reactions were carried out in THF or toluene in the presence of 2 equivalents of TBSCl, 1.25 equivalents of DIPEA, and with a substrate concentration of 0.5 or 1.0 M (see the Supplementary Information for details). See the Supplementary information for proof of product absolute stereochemistry. The yield of isolated product after purification is given (see the Supplementary Information for details). The yield shown in entry number 10 is based on GLC analysis; isolated yield in this case is 47%, owing to high compound volatility.

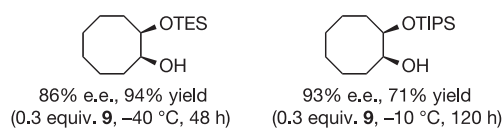


Figure 2 | Enantioselective silylations of achiral diols with different silyl chloride reagents catalysed by **9.** TES, triethylsilyl; TIPS, tri(*iso*-propyl)silyl.

(*tert*-butyldimethylsilyl chloride) and diisopropylethylamine (DIPEA; to quench the HCl byproduct). We thus established that, unlike compounds such as **5**, **7** and **8** (see Table 1), which generate racemic **3** (50:50 e.r.), silylation of **1** in the presence of **6** (a compound that resembles amino-acid-based ligands developed in these laboratories²⁰) affords **3** with measurable optical enrichment (58:42 e.r., 16% e.e.). This was encouraging because control experiments indicated that the uncatalysed reaction (to give racemic product) occurs to a small extent. When we performed the silylation at -78 °C to minimize uncatalysed silylation, we obtained **3** in 74:26 e.r. (48% e.e.). Reactions at -78 °C with **5**, **7** and **8** afforded racemic **3**.

Next, the modularity of peptidic structures was exploited through positional optimization strategies²¹; analogues of **6** were prepared and examined. These studies (Supplementary Information SI4) indicated that in the presence of 0.2 equivalents of **9**, 2 equivalents of TBSCl, and 1.25 equivalents of DIPEA, enantioselective silylation of **1** leads to the formation of **3** (see entry 1 in Table 2) in 90.5:9.5 e.r. (81% e.e.) and 54% yield. Further studies, summarized in Table 2, established that **9** promotes the enantioselective silylation of a range of achiral diols, giving rise to chiral silyl ethers in excellent enantiomeric purity ($\geq 94:6$ e.r., $\geq 88\%$ e.e.).

Because **9** can easily be structurally altered, when higher enantioselectivity is desired, modified catalysts could be investigated. For example, as shown in entry 2 of Table 2, formation of **3**, the least enantioselective silylation in Table 2, proceeds to >90% conversion in 48 h in 87% e.e. and 55% yield when 20 mol.% of a catalyst that bears an *iso*-Leu rather than a *tert*-Leu moiety (**10**) is used. This specific modification does not lead to improvement in enantioselectivity for the other transformations in Table 2; if necessary, further screening may be carried out in the case of other substrates²².

Table 3 | Effect of variations in catalyst structure on enantioselective synthesis of **3**

Entry number	Catalyst	Conversion (%)	e.r.	e.e. (%)
1	11	<5	-	-
2	12	<5	-	-
3	13	<5	-	-
4	14	<5	-	-
5	15	19	83.5:16.5	67
6	16	13	62.5:37.5	25
7	17	19	65:35	30
8	18	42	81:19	62

The reactions were carried out with 0.2 equivalents of catalyst and 1.25 equivalents DIPEA in THF at -78 °C for 24 h. The conversion was determined by GLC analysis of unpurified product mixtures.

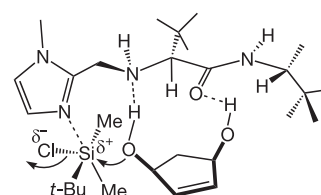


Figure 3 | Proposed transition state model for catalytic enantioselective silylation of diol **1.**

Catalytic silylations can be performed on 1,2-diols with high enantioselectivity (entries 4–11 of Table 2), and are not limited to reactions of five-membered rings: six-membered (entries 5 and 6) and medium-sized ring substrates (entries 7–9) are transformed to silyl ethers with >92% e.e. (>96:4 e.r.). As the examples in entries 10–11 of Table 2 indicate, acyclic substrates are desymmetrized to afford products of high enantiomeric purity (>95:5 e.r., >90% e.e.) and in useful yields. As the examples in Fig. 2 illustrate, other commercially available silyl chlorides, such as TESCl (triethylsilyl chloride) and TIPSCl [tri(*iso*-propyl)silyl chloride] can be used to afford the desired optically enriched compounds.

Catalyst **9**, a small molecule (formula weight, 308.5 g mol⁻¹), is prepared from commercially available materials in 60–70% overall yield by a straightforward three-step sequence (see Supplementary Information SI3 and SI4). The catalyst preparation does not require distilled solvents, expensive inert gases or special laboratory apparatus. Analytically pure catalyst can be obtained after one simple silica gel chromatography. A shortcoming of the method is that it requires relatively high catalyst loadings. However, this deficiency is tempered by the practical advantage that the chiral catalyst is easily recovered in quantitative (>99%) yield after a mild aqueous wash. The recovered catalyst can be re-used with the initial efficiency and enantioselectivity (see Supplementary Information SI5 and SI12). Catalytic silylations are simple to perform and do not require significant expertise in chemical synthesis. Reactions can be performed directly with reagents (silyl chlorides, solvents, DIPEA) purchased from popular vendors without purification, and do not require rigorous exclusion of air and moisture. A minimal amount of solvent is sufficient (0.5–1.0 M); solvent distillation is not needed.

In nearly all reactions small amounts (<3%) of bis-silylated products are isolated, but in the case of the reaction in entry 1 of Table 2, up to ~25% of the bis(silyl) product can be formed. In the latter instance, high selectivities are not due to rapid second silylation of the minor product enantiomer (kinetic resolution)¹⁹. This was established through subjection of a racemic sample of **3** to the reaction conditions. After the appropriate time (Table 2), <5% bis(silyl) compound was observed and racemic substrate was recovered.

We propose the transition state model in Fig. 3. Substrate–catalyst association by H-bonding²³ results in silylation of one enantiotopic alcohol. On the basis of principles presented by Guttmann¹³ and developed by Denmark¹⁴, by serving as a donor ligand to the Si centre, the imidazole moiety promotes re-distribution of electron density and enhancement of Si electrophilicity (increased build-up of electron density at ligands of Si). Treatment of **9** with TBSCl does not afford a salt from reaction of the imidazole with the silylchloride²⁴. Although the intermediacy of such a salt²³ cannot be ruled out, its involvement is not required. The suggested scenario is supported by the significant change in activity exhibited by catalyst candidates that are structurally modified (Table 3).

Compound **11**, bearing a less nucleophilic thiazole, is ineffective (entry 1, Table 3). The Schiff base **12** (entry 2, Table 3) is inactive; the N atom of the C=N in **12** (versus C–N in **9**) may not be able to H-bond effectively with the substrate hydroxyl group due to geometric constraints (see Fig. 3). Amide **13** (entry 3, Table 3) is inactive, probably because the H-bond acceptor has been removed from the catalyst structure. The Lewis-basic but more hindered tertiary amine **14** is inactive; H-bonding, illustrated in Fig. 3, would destabilize the proposed mode of catalysis by producing unfavourable steric interactions between the amine's methyl group and the *tert*-Leu moiety. The importance of the H-bonding is emphasized by loss of activity (19% versus 70% conversion with **9**) and lower enantioselectivity (67% e.e. versus 81% e.e. with **9**) caused by the removal of the amide carbonyl of the *tert*-Leu (**15**). Thus, the less-Lewis-basic thiocarbonyl of **16** offers low (25% versus 81% e.e.) enantioselectivity at a reduced rate (13% versus 70% conversion), and the still lower Lewis basicity of *tert*-butyl ester of **17** translates to an inefficient (19% conversion) reaction that affords **3** in 65:35 e.r. (30% e.e.). The lower selectivity and efficiency observed with *tert*-butylamide **18** points to the chiral amine group as a critical element. A study of enantioselective silylations with samples of **9** of varying enantiomeric purity indicated that there is a linear relationship²⁵ between catalyst and product e.e. (see the Supplementary Information page SI11); this suggests that substrate association and catalysis by complexes that consist of several molecules of **9** is less likely.

METHODS

General procedure for catalytic enantioselective silylation. Catalyst **9** (>98% e.e., >99:1 e.r.) and the diol substrate were weighed into a 10 × 75 mm test tube and DIPEA was added with a Gilson Pipetman. The tube's contents were dissolved in THF (or toluene), capped with a septum, and cooled to the appropriate temperature (see Table 2). In a separate vial, TBSCl was dissolved in THF (or toluene), cooled to the above temperature and added to the test tube with a Gilson Pipetman. The test tube was capped with a septum, wrapped with Teflon tape and the mixture was allowed to stir for the appropriate period of time (see Table 2). The reaction was quenched by addition of DIPEA (1.0 equivalents relative to substrate) followed by methanol (0.2 ml). The mixture was allowed to warm to 22 °C, diluted with CH₂Cl₂ (15 ml) and washed with 20 ml of a 10% aqueous solution of citric acid. The aqueous layer was washed with CH₂Cl₂ (2 × 15 ml) and the combined organic layers were dried over MgSO₄, solids were removed by filtration and the solution was concentrated in vacuum to produce the unpurified product as yellow oil. The resulting yellow oil was purified by silica gel chromatography and its enantiomeric purity was analysed by chiral GLC (Supelco Beta, or Gamma Dex 120). The catalyst was recovered in the following manner: The aqueous layer was rendered basic through treatment with an aqueous solution of 3M NaOH until pH = 12 and then it was washed with 3 × 15 ml portions of CH₂Cl₂. The combined organic layers were dried over MgSO₄, solids were removed by filtration and the solution was concentrated in vacuum to provide the recovered catalyst as a white solid (mass recovery >98%).

Received 7 March; accepted 24 July 2006.

1. Thayer, A. Chiral catalysts. *Chem. Eng. News* **83**, 40–48 (2005).
2. Kocienski, P. J. *Protecting Groups*. (Thieme, Stuttgart, 2005).
3. Jacobsen, E. N., Pfaltz, A. & Yamamoto, H. (eds) *Comprehensive Asymmetric Catalysis* (Springer, Berlin, 1999).
4. Corey, E. J. & Cheng, X.-M. *The Logic of Chemical Synthesis* (Wiley, New York, 1989).
5. Corey, E. J. & Venkateswarlu, A. Protection of hydroxyl groups as *tert*-butyldimethylsilyl derivatives. *J. Am. Chem. Soc.* **94**, 6190–6191 (1972).
6. Isobe, T., Fukuda, K., Araki, Y. & Ishikawa, T. Modified guanidines as chiral superbases: the first example of asymmetric silylation of secondary alcohols. *Chem. Commun.*, 243–244 (2001).
7. Rendler, S., Auer, G. & Oestreich, M. Kinetic resolution of chiral secondary alcohols by dehydrogenative coupling with recyclable silicon-stereogenic silanes. *Angew. Chem. Int. Edn* **44**, 7620–7624 (2005).
8. Jarvo, E. R. & Miller, S. J. Asymmetric acylation in *Comprehensive Asymmetric Catalysis* (Supplement 1) (eds Jacobsen, E. N., Pfaltz, A. & Yamamoto, H.) 189–206 (Springer, Berlin, 2004).
9. Miller, S. J., Copeland, G. T., Papaioannou, N., Horstmann, T. E. & Ruel, E. M. Kinetic resolution of alcohols catalyzed by tripeptides containing the *N*-alkylimidazole substructure. *J. Am. Chem. Soc.* **120**, 1629–1630 (1998).
10. Sculimbrene, B. & Miller, S. J. Discovery of a catalytic asymmetric phosphorylation through selection of a minimal kinase mimic: A concise total synthesis of D-myoinositol-1-phosphate. *J. Am. Chem. Soc.* **123**, 10125–10126 (2001).
11. Evans, J. W., Fierman, M. B., Miller, S. J. & Ellman, J. A. Catalytic enantioselective synthesis of sulfinate esters through the dynamic resolution of *tert*-butanesulfinyl chloride. *J. Am. Chem. Soc.* **126**, 8134–8135 (2004).
12. Myers, A. G., Hammond, M. & Wu, Y. An improved preparation of highly enantiomerically enriched (R)-(+)-4-*tert*-butyl-di-methylsiloxy-2-cyclopenten-1-one. *Tetrahedr. Lett.* **37**, 3083–3086 (1996).
13. Guttmann, V. *The Donor–Acceptor Approach to Molecular Interactions* (Plenum, New York, 1978).
14. Denmark, S. E., Beutner, G. L., Wynn, T. & Eastgate, M. D. Lewis base activation of Lewis acids: catalytic enantioselective addition of silyl ketene acetals to aldehydes. *J. Am. Chem. Soc.* **127**, 3774–3789 (2005).
15. Kelly, T. R., Zhao, C. & Bridger, G. J. A bisubstrate reaction template. *J. Am. Chem. Soc.* **111**, 3744–3745 (1989).
16. Jarvo, E. R., Copeland, G. T., Papaioannou, N., Bonitatebus, P. J. & Miller, S. J. A biomimetic approach to asymmetric acyl transfer catalysis. *J. Am. Chem. Soc.* **121**, 11638–11643 (1999).
17. Taylor, M. S. & Jacobsen, E. N. Asymmetric catalysis by chiral hydrogen-bond donors. *Angew. Chem. Int. Edn* **45**, 1520–1543 (2006).
18. Traverse, J. F., Zhao, Y., Hoveyda, A. H. & Snapper, M. L. Proline-based *N*-oxides as readily available and modular chiral catalysts. Enantioselective reactions of allyltrichlorosilane with aldehydes. *Org. Lett.* **7**, 3151–3154 (2005).
19. Vedejs, E. & Jure, M. Efficiency in nonenzymatic kinetic resolution. *Angew. Chem. Int. Edn* **44**, 3974–4001 (2005).
20. Porter, J., Traverse, J. F., Hoveyda, A. H. & Snapper, M. L. Enantioselective synthesis of arylamines through Zr-catalyzed addition of dialkylzincs to imines. Reaction development by screening of parallel libraries. *J. Am. Chem. Soc.* **123**, 984–985 (2001).
21. Cole, B. M. *et al.* Chiral catalyst discovery through ligand diversity. Ti-catalyzed enantioselective addition of TMSCN to meso epoxides. *Angew. Chem. Int. Edn* **35**, 1668–1671 (1996).
22. Hoveyda, A. H. in *Handbook of Combinatorial Chemistry* (eds Nicolaou, K. C., Hanko, R. & Hartwig, W.) 991–1016 (Wiley-VCH, Weinheim, 2002).
23. Vasbinder, M. M., Jarvo, E. R. & Miller, S. J. Incorporation of peptide isosteres into enantioselective peptide-based catalysts as mechanistic probes. *Angew. Chem. Int. Edn* **40**, 2824–2827 (2001).
24. Bassindale, A. R. & Stout, T. The interaction of electrophilic silanes (Me₂SiX, X = ClO₄, I, CF₃SO₃, Br, Cl) with nucleophiles. The nature of silylation mixtures in solution. *Tetrahedr. Lett.* **26**, 3403–3406 (1985).
25. Blackmond, D. Kinetic aspects of nonlinear effects in asymmetric catalysis. *Acc. Chem. Res.* **33**, 402–411 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to our colleague S. J. Miller for discussions and a critical reading of the manuscript. This work was supported by the United States National Institutes of Health, Institute of General Medical Sciences (grant to A.H.H. and M.L.S.). We are grateful to J. F. Traverse for experimental assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.H.H. (amir.hoveyda@bc.edu) or M.L.S. (marc.snapper@bc.edu).

Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming

K. M. Walter¹, S. A. Zimov², J. P. Chanton³, D. Verbyla⁴ & F. S. Chapin III¹

Large uncertainties in the budget of atmospheric methane, an important greenhouse gas, limit the accuracy of climate change projections^{1,2}. Thaw lakes in North Siberia are known to emit methane³, but the magnitude of these emissions remains uncertain because most methane is released through ebullition (bubbling), which is spatially and temporally variable. Here we report a new method of measuring ebullition and use it to quantify methane emissions from two thaw lakes in North Siberia. We show that ebullition accounts for 95 per cent of methane emissions from these lakes, and that methane flux from thaw lakes in our study region may be five times higher than previously estimated³. Extrapolation of these fluxes indicates that thaw lakes in North Siberia emit 3.8 teragrams of methane per year, which increases present estimates of methane emissions from northern wetlands (<6–40 teragrams per year; refs 1, 2, 4–6) by between 10 and 63 per cent. We find that thawing permafrost along lake margins accounts for most of the methane released from the lakes, and estimate that an expansion of thaw lakes between 1974 and 2000, which was concurrent with regional warming, increased methane emissions in our study region by 58 per cent. Furthermore, the Pleistocene age (35,260–42,900 years) of methane emitted from hotspots along thawing lake margins indicates that this positive feedback to climate warming has led to the release of old carbon stocks previously stored in permafrost.

Understanding the role of North Siberian thaw lakes in the global atmospheric methane (CH₄) budget is important because the concentration of atmospheric CH₄, a potent greenhouse gas, is highest at 65° to 70° N (refs 1, 7), has risen during recent decades^{1,8}, and exhibits an inexplicably large seasonal amplitude at high northern latitudes with bimodal maxima in winter and spring⁷—times of year when wetlands are frozen, but unfrozen lake sediments actively produce and emit CH₄. This study shows that ebullition from Siberian thaw lakes is a large and increasing source of atmospheric CH₄ as Siberian thaw lakes continue to expand.

Thaw lakes comprise 90% of the lakes in the Russian permafrost zone⁹. Many North Siberian lakes differ from lakes in America and Europe because they are underlain by yedoma, an organic-rich (~2% carbon by mass) Pleistocene-age loess permafrost with ice content of 50–90% by volume^{3,10,11}. Thermokarst erosion occurs along lake margins when massive, subsurface ground ice wedges melt, causing the ground surface to subside. Labile organic matter from permafrost erodes into anaerobic lake bottoms when yedoma thaws, enhancing methane production and emission³. In this study we used remote sensing, aerial surveys and year-round, continuous measurements of CH₄ flux to (1) quantify CH₄ emissions from North Siberian lakes, paying particular attention to the role of ebullition, and (2) document the role of thermokarst erosion as a landscape process that fuels CH₄ production and feeds back to climate warming.

Ebullition is often the dominant pathway of CH₄ release from aquatic ecosystems^{12,13}, yet it is seldom measured owing to its extreme temporal and spatial patchiness. Siberian lakes provide a unique opportunity to assess ebullition flux more accurately. As ice forms in autumn, bubbles released from lake sediments are frozen in place, allowing accurate mapping of ebullition 'point sources' across lake surfaces (Fig. 1). We used random and selective placement of bubble traps under water or ice to make continuous measurements of 'background' (measured with randomly placed traps), 'point-source' (moderately high bubble flux from discrete points in lake sediments) and 'hotspot' (open holes in ice due to extremely high bubble flux from discrete points in lake sediments) fluxes from April 2003 through May 2004 on two intensively studied lakes in North Siberia.

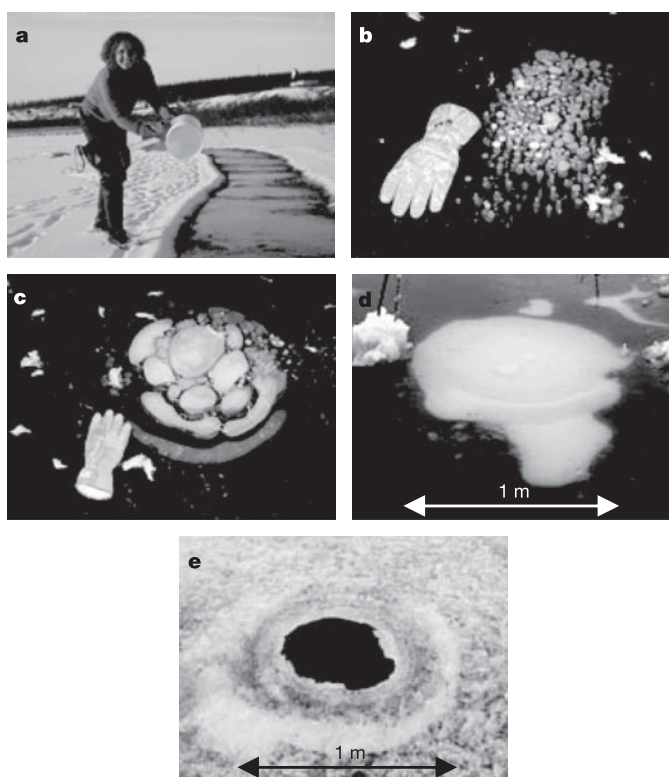


Figure 1 | Photographs of CH₄ bubbles trapped in lake ice represent point sources of bubbling. The distribution of ebullition point sources was measured along transects on lake ice (a), recognizing four distinct categories of bubble clusters with associated CH₄ emission rates (Supplementary Information): b, *kotenok*; c, *koshka*; d, *kotara*; e, *hotspot*.

¹Institute of Arctic Biology, University of Alaska Fairbanks, Alaska 99775, USA. ²Northeast Science Station, Cherskii 678830, Russia. ³Department of Oceanography, Florida State University, Florida 32306, USA. ⁴Forest Science Department, University of Alaska Fairbanks, Alaska 99775, USA.

Table 1 | Summary of CH₄ flux components and isotopes in bubbles from North Siberian thermokarst lakes

Season 2003–04	Flux component	Whole lake (g CH ₄ m ⁻² yr ⁻¹)	Percentage of annual flux	Thermokarst margins (g CH ₄ m ⁻² yr ⁻¹)	¹⁴ CH ₄ age (years)	¹⁴ CH ₄ (% modern carbon)	δ ¹³ C of CH ₄ (‰)	δD of CH ₄ (‰)
Summer (open water June–September)	Background ebullition	5.7 ± 2.0	22 ± 6	27.7 ± 14	1,345 to 5,350	51–85	–57.7 to –73.7	–420 to –338
	Point source	6.2 ± 0.7	25 ± 5	34.7 ± 0.7	5,590 to 12,285	22–50	–62.6 to –82.8	–388 to –376
	Hotspots	1.0 ± 0.7	4 ± 2	5.4 ± 3.9	35,570 to 39,120	0.77–1.2	–78.5 to –80.4	–401 to –392
	Molecular diffusion	1.3 ± 0.1	5 ± 1	1.3 ± 0.1				
Winter (ice cover October–May)	Background ebullition	0.5 ± 0	2 ± 0	0.7 ± 0				
	Point source	8.3 ± 0.9	34 ± 7	46.9 ± 0.9	22,050	6.4	–76.0 to –78.3	–394 to –379
	Hotspots	2.0 ± 1.3	8 ± 5	11.1 ± 7.7	35,260 to 42,900	0.49–1.2	–71.7 to –80.4	–420 to –388
Total		24.9 ± 2.3	100	127.8 ± 24.0	16,520†			

† Sum of the mean radiocarbon age of each flux component weighted by each component's proportion of total annual emissions. Forty-four per cent of annual CH₄ emissions from Siberian lakes occurs in winter and spring when CH₄ trapped as bubble clusters in lake ice is released to the atmosphere during high temperature anomalies³, overflow events when lake water is forced through lenses of ice clusters and open holes after heavy snowfall, and springtime thaw of lake ice¹⁷. Error estimates are half of the absolute difference between measurements for two intensive study lakes. Sources of variability are shown in Supplementary Table 1.

We extrapolated fluxes regionally using the distribution of hotspots in aerial photographs.

Ebullition comprised 95 ± 1% of the CH₄ flux from the intensively studied lakes, with molecular diffusion contributing the remainder (Table 1). Emergent plants were absent as a conduit in our lakes. Point sources comprised the largest proportion (59 ± 12%) of total CH₄ emissions. Ebullition rates were highest in traps selectively placed over hotspots along thermokarst margins of lakes (Fig. 2), with hotspots <1% of lake area but contributing 11 ± 7% of the total CH₄ flux. The maximum rate of 18,835 mg of CH₄ per day per hotspot (>30 litres per day per hotspot) exceeds most published CH₄ flux rates for lakes and wetlands by several orders of magnitude^{4,5,12–17}.

Hotspots were visible as 'black holes' in lake ice along thermokarst margins in ground and aerial surveys of lakes (Supplementary Information). Half of the 60 lakes surveyed in our study region (including our two intensively studied lakes) had modest thermo-

karst erosion, marked by banks with gentle slopes, stable vegetation and hotspots distributed primarily in a 15-m-wide belt along thaw margins. The remaining lakes had more intense thermokarst erosion, with steep muddy banks, occasional exposed ice wedges, and a >30-m-wide belt of hotspot emissions at the lake margins.

When background, point-source, and hotspot ebullition fluxes were aggregated to a whole-lake basis, the annual flux was 24.9 ± 2.3 g CH₄ m⁻² yr⁻¹ from the intensively studied lakes. Our estimate is conservative because it neglects other extreme expulsions of CH₄ from lakes, which were not measured because of their rarity (Supplementary Information). Fluxes from the 15-m-wide band of active thermokarst on the eroding edge of lakes (15.8% of lake area, 128 ± 24 g CH₄ m⁻² yr⁻¹) accounted for 79% of total lake emissions. In the other 50% of the region's lakes, where we extrapolated the thermokarst margin flux to the >30-m-wide band in which hotspots were observed (31.6% of lake area), we estimate that thermokarst margins accounted for 90% of the lake-averaged annual flux of 43.7 ± 3.2 g CH₄ m⁻² yr⁻¹.

These emission rates exceed those reported in previous work on the same lakes (6.8 g CH₄ m⁻² yr⁻¹; ref. 3), other high-latitude lakes and ponds (0.5–11.3 g CH₄ m⁻² yr⁻¹; refs 17–19), and North Siberian wetlands (0.9–14.3 g CH₄ m⁻² yr⁻¹; refs 20, 21) and 6.4 ± 3.4 g CH₄ m⁻² yr⁻¹ (*n* = 8 chambers measured biweekly from June–October in 2003; this study).

Our lake estimate probably exceeds others not only because of the prevalence of thermokarst activity in North Siberia, but also because our stratified sampling design with continuous measurements and mapping point sources and hotspots accounted for the patchiness of bubbling. The few studies that previously measured ebullition estimated fluxes from a few randomly placed bubble traps that were periodically deployed^{13,12,13,22}. In our study, random placement of up to 14 traps per lake with continuous, year-round monitoring revealed that this 'background' bubbling accounted for only 24 ± 6% of the total emissions. To our knowledge, this is the first study to quantify the spatial and temporal patchiness of ebullition in any lake, thereby reducing uncertainty in a highly sporadic mode of methane emission.

We extrapolated our lake flux estimate to the yedoma region of North Siberia using an estimated lake area for the region of 11% as measured in our Geographical Information Systems (GIS) study—which is conservative given the range of lake area estimates for this region (8.5 to >30% of the yedoma region covered by lakes; refs 3, 9, 23)—yielding a regional flux of 3.8 Tg CH₄ yr⁻¹. Using an independent mass-balance approach based on organic carbon loss from yedoma beneath thaw lakes, we estimated that decomposition of yedoma beneath Siberian lakes contributed on average 3.7 Tg CH₄ yr⁻¹ to the atmosphere throughout the Holocene. Given the uncertainties associated with both flux estimates (Supplementary Information), their strong correspondence supports our conclusion that North Siberian lakes are a globally significant source of atmospheric CH₄.

Adding our results of ebullition from North Siberian lakes

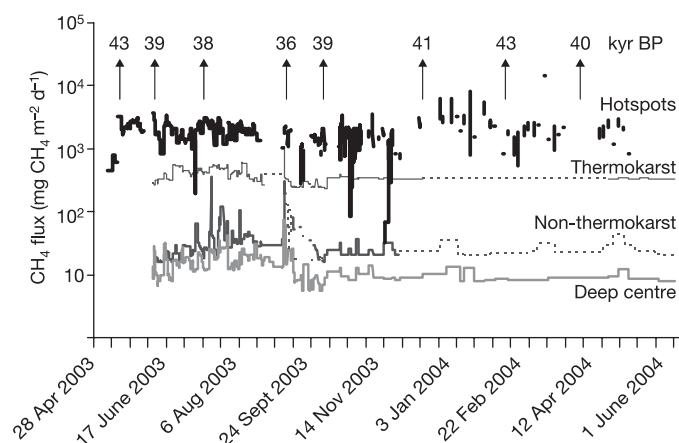


Figure 2 | Average ebullition rates for representative areas of lakes and hotspots. Ebullition rates along thermokarst margins (15.8 ± 2.2% of lake area) were an order of magnitude greater than ebullition rates along non-thermokarst margins (4.4 ± 1.4% of lake area) and at various depths throughout the lake centre (79.8 ± 0.8% of lake area). Hotspots (<1% of lake area) are presented here separately, and units of hotspot emission are distinct: mg CH₄ day⁻¹ spot⁻¹. Ebullition from a hotspot occurs through tiny bubble tubes (<2 cm diameter) at the sediment surface, but represents CH₄ produced in a larger volume of lake sediments. Hotspot emissions are extrapolated by area on the basis of the number of hotspots per unit area of lake surface (Table 1). Breaks in the hotspot line are periods of no data (Supplementary Information). Numbers above hotspot data indicate the radiocarbon ¹⁴C age (kyr BP) of CH₄ bubbles collected from a particular hotspot repeatedly throughout the study period. We interpolated fluxes (dashed lines) if fewer than 25% of the total traps were operating owing to temporary mechanical failures.

(3.8 Tg yr^{-1}), which have not yet been included in global estimates of CH_4 emissions from wetlands^{2,15,16,24}, to a range of estimates for northern wetland flux ($<6\text{--}40 \text{ Tg CH}_4 \text{ yr}^{-1}$; refs 1, 2, 4–6) increases current estimates of northern wetland CH_4 emissions by 10–63%. Because yedoma lakes are only a fraction of northern lakes, ebullition measurements in other lake regions would probably further increase CH_4 emission estimates.

In the zone of continuous permafrost, which comprises about half of arctic permafrost²⁵, warming²⁶ and degradation of permafrost have caused thaw lakes to grow in number and size during recent decades²⁷. Using GIS analysis of 1974 Multispectral Scanner (MSS) and 2000 Enhanced Thematic Mapper (ETM+) Landsat imagery, we measured a 14.7% increase in lake area (from 9.6% to 11% lakes) for a 12,000 km^2 territory including our study region along the Kolyma River near Cherskii, Russia (Fig. 3). This is similar to the 12% increase in lake area observed in continuous permafrost zones of West Siberia during the same period²⁷. Applying CH_4 emission rates associated with thaw margins to areas of lake expansion based on observations of hotspot distributions (see Methods), we estimated that the 14.7% increase in lake area resulted in a 58% increase in lake CH_4 emissions, or $1.4 \text{ Tg CH}_4 \text{ yr}^{-1}$ between 1974 ($2.4 \text{ Tg CH}_4 \text{ yr}^{-1}$) and 2000 ($3.8 \text{ Tg CH}_4 \text{ yr}^{-1}$) if extrapolated regionally. Regional warming observed during the study period²⁸ of 1974–2004 (Fig. 3) is consistent with lake expansion that contributed to rising atmospheric CH_4 concentration and global temperature. This is, however, not a simple feedback loop because many processes contribute to thermokarst lake expansion and climate warming over a range of timescales.

Isotopic analyses revealed that methane collected from ebullition

in Siberian thaw lakes during 2001–2004 is of biogenic origin ($\delta^{13}\text{C}$ of $\text{CH}_4 = -58\text{‰}$ to -83‰ , δD of $\text{CH}_4 = -338\text{‰}$ to -420‰) (Table 1). Hotspot $^{14}\text{CH}_4$, ranging from 35,260–42,900 yr (0.49–1.2% modern carbon), reflects the importance of recently thawed Pleistocene-age permafrost soils as a source of labile organic matter for methanogenesis deep within the thaw bulb of lakes (Fig. 2). Non-hotspot sources of bubbling had younger radiocarbon ages, 1,385–22,050 yr (6.4–85% modern carbon), indicating a greater influence of Holocene organic sources for methane production. Weighting the radiocarbon signature of point-source types by their abundance in lakes yielded an average annual $^{14}\text{CH}_4$ age of 16,520 yr, signifying that Pleistocene sediments deposited 20,000–40,000 yr ago constitute 41–83% of the CH_4 emitted from Siberian lakes. The radiocarbon age of CH_4 sampled from our North Siberian thaw lakes is distinct from other arctic lakes and wetlands, including the East Siberian alasses situated on sandy soils whose ^{14}C of CH_4 value is 95–123% modern carbon²⁹. Therefore the unique ^{14}C -depleted CH_4 signature of ebullition from yedoma lakes should be considered in future studies that estimate the fraction of total atmospheric CH_4 emissions derived from fossil fuels.

In conclusion, we have shown that North Siberian lakes are a significantly larger source of atmospheric CH_4 than previously recognized. Emissions are dominated by ebullition, a mode of emission that we have quantified using the new technique of mapping bubbling point sources. This CH_4 source is largely fuelled by thermokarst, and we have linked the expansion of thaw lakes during recent decades with a 58% increase in lake CH_4 emissions, demonstrating a new feedback to climate warming. Though the recent increase in flux due to lake expansion is modest relative to

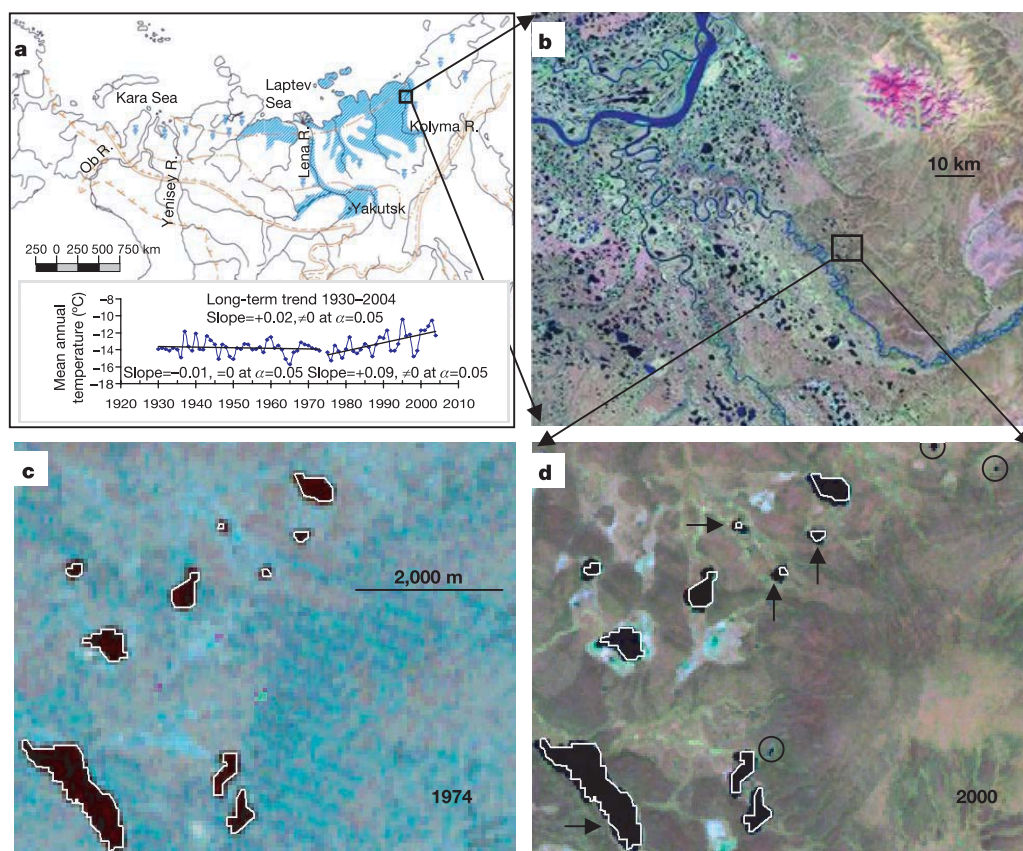


Figure 3 | Thermokarst lake expansion. The schematic in **a** shows yedoma distribution (blue) in North Siberia and the mean annual air temperature for the half-degree-resolution grid cell that includes our study area (box in **a** indicates location of image in **b**), interpolated from climate station records from 1930–2004 (ref. 30). Blue shading shows continuous yedoma distribution (adapted from ref. 32); blue triangles show sporadic yedoma

distribution. α , significance level. The Landsat image in **b** ($\sim 12,000 \text{ km}^2$) represents the region of our GIS lake change analysis from 1974 (**c**) to 2000 (**d**). The box in **b** indicates location of images in **c** and **d**. The shorelines from 1974 are delineated as white polygons, while some areas of expansion of thaw lakes are indicated by arrows and formation of new lakes by circles on the 2000 image in **d**.

anthropogenic emissions^{1,2,7}, the ~500 Gt of labile Pleistocene-aged C in ice-rich yedoma permafrost¹¹ could greatly intensify the positive feedback to high-latitude warming by releasing tens of thousands of teragrams of CH₄ through ebullition from thermokarst lakes if northeast Siberia continues to warm in the future, as projected²⁶.

METHODS

All numbers following a $\pm$ sign are standard deviations unless otherwise noted. Equations were developed to extrapolate flux measurements to whole lake and regional emission estimates (Supplementary Information).

Ebullition measured by bubble traps. We captured ebullition continuously using umbrella-shaped floating bubble traps (~1-m diameter, 25 traps randomly placed for background ebullition, 16 traps fixed over point sources and hotspots of bubbling) with daily measurements of bubble volume from April 26, 2003 through June 1, 2004 at two lakes (maximum depths 16.5 m and 11 m, with areas 0.11 km² and 0.06 km²) near the Northeast Science Station in Cherskii (68° 45' 36" N, 161° 20' 34" E). Gas concentrations and isotopes were also measured (Supplementary Information). To verify that our intensively studied lakes were regionally representative of thermokarst lakes on yedoma territory, we made ground surveys of 35 other lakes and took aerial photographs of 60 lakes in Northern Siberia.

Stable isotope compositions are expressed in δ (‰) = $10^3 \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right)$, where R is ¹³C/¹²C or D/H and standards refer to the Vienna Pee Dee belemnite (VPDB) and Vienna standard mean ocean water (VSMOW), respectively. The analytical errors of the stable isotopic analyses are ± 0.1 ‰ for $\delta^{13}\text{C}$ and ± 0.1 ‰ for δD . We express radiocarbon data as the percentage of modern carbon, $\left(\frac{^{14}\text{C}/^{12}\text{C}}{^{14}\text{C}/^{12}\text{C}} \right)_{\text{sample}} / \left(\frac{^{14}\text{C}/^{12}\text{C}}{^{14}\text{C}/^{12}\text{C}} \right)_{\text{standard}} \times 100$, which is the percentage of ¹⁴C/¹²C ratio normalized to $\delta^{13}\text{C} = -25$ ‰ and decay-corrected relative to that of an oxalic standard in 1950 (ref. 30).

Mapping bubble clusters in lake ice as 'point sources' of ebullition. We removed snow from lake surfaces along six 50 m $\times$ 1 m transects per lake (Fig. 1) in early winter to count the number and type of bubble clusters in lake ice that represent 'point sources' of ebullition. We categorized four types of bubble patterns in lake ice and measured winter CH₄ flux rates (mg CH₄ day⁻¹ per point) associated with each 'point-source' category and hotspots using floating traps under the ice (Supplementary Information). Average emissions from point sources and hotspots were extrapolated to entire lakes based on transects of bubble-cluster and hotspot densities in ice over shallow, medium and deep water depths in thermokarst and non-thermokarst areas of lakes. We include the estimated flux from point sources in addition to the randomly placed bubble trap flux (background ebullition) because the probability of capturing point sources and hotspots in randomly placed traps was 0.001% and there was no overlap between values of point-source fluxes and values of background ebullition.

Mass-balance approach to Holocene lake emissions. To assess the plausibility of our CH₄ flux estimate of 3.8 Tg CH₄ yr⁻¹ we used an independent mass-balance approach based on Holocene carbon loss associated with thermokarst lake development. The mean carbon content of decomposed yedoma beneath two former thermokarst lakes (18.0 ± 1.4 kg C m⁻³, standard error of 15 samples) is 33% less than in undisturbed yedoma (26.8 ± 1.5 kg C m⁻³, standard error of 54 samples). We assumed half of this carbon was converted to CH₄, on the basis of the stoichiometry of methane production from cellulose degradation³¹. Applying this 16.5% carbon loss due to methanogenesis beneath lakes to yedoma territory of North Siberia (1×10^6 km², 25 m average thickness, 50% ice content)^{3,10,23}, which has been degraded by migratory lakes whose scars cover 50% of the region²³, yields a release of 28 Gt C from CH₄ during the Holocene (10,000 yr) or on average, 3.7 Tg CH₄ yr⁻¹. This estimate is conservative because it does not include CH₄ derived from Holocene-age lake deposits.

Future CH₄ release via ebullition from expanding thaw lakes in the yedoma region is estimated to be of the order of tens of thousands of teragrams based on the amount of labile carbon remaining in yedoma permafrost (500 Gt; ref. 11) and the proportion of carbon converted to CH₄ during methanogenesis (16.5%).

Estimating CH₄ fluxes from expanding lakes. The years of the images analysed had temperature and precipitation values typical of the twentieth century for our study region²⁸. We assumed flux rates for lakes in 1974 were the same as current rates of our two intensive study lakes (24.9 g CH₄ m⁻² yr⁻¹). Applying this modest rate to 50% of the region's lakes in 2000 and higher rates (43.7 g CH₄ m⁻² yr⁻¹) to the other half of the region's lakes, where intense erosion has led to the observed 14.7% increase in lake area, resulted in a 58% increase in lake CH₄ emissions. Extrapolating CH₄ fluxes associated with lake expansion to the yedoma region yields an estimated increase in lake emissions of 1.4 Tg CH₄ yr⁻¹ between 1974 (2.4 Tg CH₄ yr⁻¹) and 2000 (3.8 Tg CH₄ yr⁻¹).

Received 5 December 2005; accepted 3 July 2006.

- Intergovernmental Panel on Climate Change (IPCC), *The Scientific Basis* (Cambridge Univ. Press, New York, 2001).
- Mikaloff Fletcher, S. E., Tans, P. P., Bruhwiler, L. M., Miller, J. B. & Heimann, M. CH₄ source estimated from atmospheric observations of CH₄ and its ¹³C/¹²C isotopic ratios: 1. Inverse modelling of source processes. *Glob. Biogeochem. Cycles* **18**, GB4004, doi:10.1029/2004GB002223 (2004).
- Zimov, S. A. et al. North Siberian lakes: A methane source fuelled by Pleistocene carbon. *Science* **277**, 800–802 (1997).
- Roulet, N. T. et al. Role of Hudson Bay lowland as a source of atmospheric methane. *J. Geophys. Res.* **99**, 1439–1454 (1994).
- Reeburgh, W. S. et al. A CH₄ emission estimate for the Kuparuk River basin, Alaska. *J. Geophys. Res.* **103**, 29005–29013 (1998).
- Worthy, D. E., Levin, J. I., Hopper, F., Ernst, M. K. & Trivett, N. B. A. Evidence for a link between climate and northern wetland methane emissions. *J. Geophys. Res.* **105**, 4031–4038 (2000).
- Fung, I. et al. Three-dimensional model synthesis of the global methane cycle. *J. Geophys. Res.* **96**, 13033–13065 (1991).
- Slugovkenko, E. J., Masarie, K. A., Lang, P. M. & Tans, P. P. Continuing decline in the growth rate of the atmospheric methane burden. *Nature* **393**, 447–450 (1998).
- Mostakhov, S. E. in *Mezhdunarodnaia konferentsiia po merzlotovedeniui, Podzemnye vody kriolitofery* (ed. Tolstikhin, N. I.) (Iakustkoe Knizhnoe Izdatel'stvo, Iakutsk, Russia, 1973).
- Romanovskii, N. N. et al. Thermokarst and land-ocean interactions, Laptev Sea Region, Russia. *Perm. Periglac. Process.* **11**, 137–152 (2002).
- Zimov, S. A., Schuur, E. A. G. & Chapin, F. S. III Permafrost and the global carbon budget. *Science* **312**, 1612–1613 (2006).
- Keller, M. & Stallard, R. F. Methane emission by bubbling from Gatun Lake, Panama. *J. Geophys. Res.* **99**, 8307–8319 (1994).
- Casper, P., Maberly, S. C., Hall, G. H. & Finlay, B. J. Fluxes of methane and carbon dioxide from a small productive lake to the atmosphere. *Biogeochemistry* **49**, 1–19 (2000).
- Kling, G., Kipphut, G. W. & Miller, M. C. The flux of CO₂ and CH₄ from lakes and rivers in arctic Alaska. *Hydrobiologia* **240**, 23–36 (1992).
- Mathews, E. & Fung, I. Methane emission from natural wetlands: Global distribution, area, and environmental characteristics of sources. *Glob. Biogeochem. Cycles* **1**, 61–86 (1987).
- Aselmann, I. & Crutzen, P. J. Global distribution of natural freshwater wetlands, rice paddies, their net primary productivity, seasonality, and possible methane emissions. *J. Atmos. Chem.* **8**, 307–358 (1989).
- Phelps, A. R., Peterson, K. M. & Jeffries, M. O. Methane efflux from high-latitude lakes during spring ice melt. *J. Geophys. Res.* **103**, 29029–29036 (1998).
- Bartlett, K. B., Crill, P. M., Sass, R. L., Harriss, R. C. & Dise, N. B. Methane emissions from tundra environments in the Yukon-Kuskokwim Delta, Alaska. *J. Geophys. Res.* **97**, 16645–16660 (1992).
- Roulet, N. T., Crill, P. M., Comer, N. T., Dove, A. & Boubonniere, R. A. CO₂ and CH₄ flux between a boreal beaver pond and the atmosphere. *J. Geophys. Res.* **102**, 29313–29319 (1997).
- Nakayama, T. Estimation of Methane Emission from Natural Wetlands in Siberian Permafrost Area. 1–100, PhD dissertation, Hokkaido Univ. (1995).
- Corradi, C., Kolle, O., Walter, K., Zimov, S. A. & Schulze, E.-D. Carbon dioxide and methane exchange of a north-east Siberian tussock tundra. *Glob. Change Biol.* **11**, 1910–1925 (2005).
- Bastviken, D., Cole, J., Pace, M. & Tranvik, L. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Glob. Biogeochem. Cycles* **18**, GB3010, doi:10.1029/2004GB002238 (2004).
- Czudek, T. & Demek, J. Thermokarst in Siberia and its influence on the development of lowland relief. *Quat. Res.* **1**, 103–120 (1970).
- Zhuang, Q. et al. Methane fluxes between terrestrial ecosystems and the atmosphere at northern high latitudes during the past century: A retrospective analysis with a process-based biogeochemistry model. *Glob. Biogeochem. Cycles* **18**, GB3010, doi:10.1029/2004GB002239 (2004).
- Brown, J., Ferrians, O. J. Jr, Heginbottom, J. A. & Melnikov, E. S. *Circum-Arctic Map of Permafrost and Ground-Ice Conditions* 1:10,000,000 (US Geological Survey Circum-Pacific Map CP-45, Reston, Virginia, 1997).
- Sazonova, T. S., Romanovsky, V. E., Walsh, J. E. & Sergueev, D. O. Permafrost dynamics in the 20th and 21st centuries along the East Siberian transect. *J. Geophys. Res.* **109**, doi:10.1029/2003JD003680 (2004).
- Smith, L. C., Sheng, Y., MacDonald, G. M. & Hinzman, L. D. Disappearing arctic lakes. *Science* **308**, 1429 (2005).
- Willmott, C. J. & Matsuura, K. *Arctic Terrestrial Air Temperature and Precipitation Monthly and Annual Time Series (1930–2000)* Version 1 (Univ. Delaware); (<http://climate.geog.udel.edu/~climate/>) or (<http://rims.unh.edu/data/data.cgi>).
- Nakagawa, F., Yoshida, N., Nojiri, Y. & Makarov, V. N. Production of methane from alasses in eastern Siberia: Implications from its ¹⁴C and stable isotope compositions. *Global Biogeochem. Cycles* **16**, doi:10.1029/2000GB001384 (2002).

30. Stuiver, M. M. & Polach, H. Reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
31. Conrad, R., Klose, M. & Clause, P. Pathway of CH_4 formation in anoxic rice field soil and rice roots determined by ^{13}C -stable isotope fractionation. *Chemosphere* **47**, 797–806 (2002).
32. Romanovsky, N. N. *Osnovy Kriogeneza Litosfery* (Izdatelstvo Moscow State University, Moscow, 1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. P. Davidov and D. A. Draluk for contributions to the research; D. Nidzgorski and E. Carr for field assistance; the Northeast Science Station for logistic support; K. Dutta, T. Schuur and the University of Florida for helping to prepare the radiocarbon targets; S. Marchenko and V. Romanovskii for assistance in analysing Cherskii climate records; and E. Dlugokencky, J. Hobbie, T. Schuur, D. Valentine, and M. S. Bret-Harte for constructive reviews. Research funding was provided by the National Science

Foundation through the Russian-American Initiative on Shelf-Land Environments of the Arctic (RAISE) of the Arctic System Science programme (ARCSS) and Polar Programs, the Environmental Protection Agency STAR Fellowship programme, and the NASA Earth System Science Fellowship programme.

Author Contributions K.M.W. was the primary investigator in this study, responsible for study design, field work, laboratory measurements, GIS and isotope analyses, data analysis, interpretation and writing. S.A.Z. and F.S.C. contributed to project planning, experimental work, and interpretation of results. J.P.C. assisted with isotope analyses and interpretation. D.V. contributed to the study design and interpretation of the GIS analysis of lake area change. All co-authors provided constructive comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.M.W. (ftkmw1@uaf.edu).

LETTERS

Magma heating by decompression-driven crystallization beneath andesite volcanoes

Jon Blundy¹, Kathy Cashman² & Madeleine Humphreys¹

Explosive volcanic eruptions are driven by exsolution of H₂O-rich vapour from silicic magma¹. Eruption dynamics involve a complex interplay between nucleation and growth of vapour bubbles and crystallization, generating highly nonlinear variation in the physical properties of magma as it ascends beneath a volcano². This makes explosive volcanism difficult to model and, ultimately, to predict. A key unknown is the temperature variation in magma rising through the sub-volcanic system, as it loses gas and crystallizes *en route*³. Thermodynamic modelling of magma that degasses, but does not crystallize, indicates that both cooling and heating are possible⁴. Hitherto it has not been possible to evaluate such alternatives because of the difficulty of tracking temperature variations in moving magma several kilometres below the surface. Here we extend recent work on glassy melt inclusions trapped in plagioclase crystals⁵ to develop a method for tracking pressure–temperature–crystallinity paths in magma beneath two active andesite volcanoes. We use dissolved H₂O in melt inclusions to constrain the pressure of H₂O at the time an inclusion became sealed, incompatible trace element concentrations to calculate the corresponding magma crystallinity and plagioclase–melt geothermometry to determine the temperature. These data are allied to ilmenite–magnetite geothermometry to show that the temperature of ascending magma increases by up to 100 °C, owing to the release of latent heat of crystallization. This heating can account for several common textural features of andesitic magmas, which might otherwise be erroneously attributed to pre-eruptive magma mixing.

Plagioclase-hosted melt inclusions from the 1980–86 eruption of Mount St Helens (Washington, USA) and the 2001–04 eruption of Shiveluch (Kamchatka, Russia) display a wide variety of shapes and textures (Fig. 1). The vast majority of inclusions are glassy rhyolites without daughter microlite crystals. There is compelling chemical and textural evidence⁵ that, subsequent to their entrapment, melt inclusions maintained some physical connectivity with the matrix melt, such that melt inclusions and matrix melt evolve in chemical harmony up until the point that the inclusions become physically isolated from the matrix. The predominant crystallizing phase at both volcanoes is plagioclase. Consequently, all inclusions show some evidence of precipitation of this mineral around their walls (Fig. 1c, d). In some cases, plagioclase precipitation proceeds to the extent that the inclusion is almost completely filled-in, leading to a pseudomorphic texture (Fig. 1c).

For those inclusions that have not been completely infilled, the relationship between the chemistry of the inclusion glass and the composition of the immediately adjacent plagioclase in-fill (for example, Fig. 1d) contains information about magmatic conditions at the point of chemical isolation. This is either the point at which the connection between inclusion and matrix became occluded by plagioclase growth, or that at which chemical diffusion along the

connection ceased to be viable on the relevant timescale. In essence, chemical equilibrium between inclusion and plagioclase provides a ‘snapshot’ of the overall ascent path of the magma in which the crystal is found⁵. We analysed 99 melt inclusions and their adjacent plagioclases from Mount St Helens and 50 from Shiveluch for H₂O, major elements and trace elements using ion- and electron-microprobes (see Methods and Table 1).

Before applying any geothermometer or geobarometer, it is

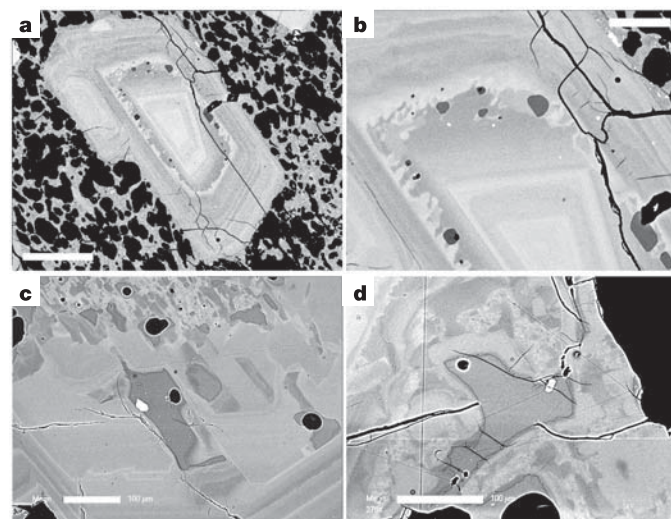


Figure 1 | Back-scattered electron micrographs of plagioclase-hosted melt inclusions from Mount St Helens. Variations in grey-scale in plagioclase denote compositional zoning from An-poor (dark) to An-rich (bright). The dark grey patches within plagioclase are melt inclusions; darker shades of grey denote higher dissolved H₂O. Black denotes cracks or gas bubbles. **a**, Melt inclusions preserved within a single resorption zone; sample was erupted on 7 August 1980. Note increase in An (lighter) on the rimward side of the resorption zone. **b**, Detail of **a**, showing presence of tiny vapour bubbles (black) in some inclusions. **c**, Melt inclusions within spongy plagioclase core; sample was erupted on 27 December 1980. Melt inclusions are surrounded by plagioclase, which itself crystallized from melt inclusions, eventually creating a chequerboard pattern of plagioclase pseudomorphs. Note oscillatory zoning in the phenocryst rim, in which wavy resorption horizons are overgrown by more An-rich plagioclase. Large inclusion in centre of the image contains tiny daughter microlites (white) of oxides and pyroxene, which also occur in the groundmass of this sample. **d**, Late-stage in-growth of plagioclase into melt inclusion; sample was erupted on 16 October 1980. As in **c**, note patchy texture produced by in-fill of the melt inclusion network. Small embayments of plagioclase into the melt inclusion, such as those just above the scale bar, are of the type used for plagioclase–melt thermometry. Scale bars 100 µm (**a**, **c**, **d**) and 25 µm (**b**).

¹Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK. ²Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403-1272, USA.

imperative to ensure that it affords adequate accuracy and precision. In the case of converting H_2O in melt inclusions to the partial pressure of H_2O (p_{H_2O}), robust solubility models are available⁶. To determine crystallinity, we used ion-microprobe analyses of the trace elements Rb and Zr, which have bulk solid–melt partition coefficients <0.05 . Crystallinity is then calculated relative to the average Rb or Zr content of the whole rock, which varies little between eruptions at each volcano. In the case of plagioclase–melt thermometry, we evaluated a recent thermometer⁷ against a set of 145 plagioclase–melt pairs from published experiments^{8–16} not used in the original calibration. These data involve hydrous rhyolitic melt at temperatures of 750–995 °C and total pressures (P_{tot}) of 15–313 MPa. The thermometer gives an absolute average deviation of ± 19 °C for this data set, which is an estimate of its accuracy. The precision of the calculated p_{H_2O} and temperature, based on propagation of analytical uncertainties, is ± 7 MPa and ± 10 °C, respectively. The p_{H_2O} can be roughly equated to P_{tot} (and hence to depth of crystallization), because the partial pressures of other dissolved volatile species (for example, CO_2 , SO_2 , H_2S , Cl and F) in these melt inclusions are relatively low⁵. For example, a small subset of bubble-free melt inclusions at Mount St Helens was measured for CO_2 , again by ion-probe. The maximum CO_2 measured is 400 p.p.m. at 6 wt% H_2O , consistent with an additional partial pressure of CO_2 of ~ 60 MPa; lower CO_2 contents of melt inclusions with <6 wt% H_2O makes the pressure correction less.

Calculated crystallinities, p_{H_2O} and temperatures for both volcanoes (Fig. 2) show a clear increase in crystallinity and temperature with decreasing p_{H_2O} . At Mount St Helens, the explosively erupted, microlite-free plinian magma of 18 May 1980 consistently records the highest p_{H_2O} , in keeping with rapid extraction of these magmas from a sub-volcanic reservoir 7 km or more below the surface¹⁷. All subsequent eruptions, in addition to pre-plinian samples derived from the shallow cryptodome, show a wide spectrum of p_{H_2O} and

temperature consistent with slower ascent and less explosive eruption of magmas stored pre-eruptively over a wide depth range¹⁷. The increase in crystallinity (Fig. 2a, d) is an expected consequence of the increase in liquidus temperature with decreasing p_{H_2O} for H_2O -saturated liquids¹⁸. The increase in temperature (Fig. 2b, e), by up to 100 °C, is at first surprising, given that it accompanies an increase in crystallinity of up to 40 wt%. For the case of crystal-free, H_2O -saturated rhyolite melt undergoing isenthalpic ascent—a reasonable scenario for slowly-ascending samples—thermodynamic calculations suggest temperature increases of the order 25 °C for a 200 MPa pressure drop⁴. The rather greater temperature increase that we observe (Fig. 2a, d) indicates that the additional contribution of latent heat to the thermal budget is significant. Using data for plagioclase crystallization from anhydrous melts, Couch *et al.*¹⁹ calculate a temperature rise of 2.3 °C per 1% of plagioclase crystallization. A similar calculation for crystallization of plagioclase + orthopyroxene + magnetite in the proportions observed at Shiveluch yields 3.2 °C per 1% crystallized. These values are entirely consistent with our data (Fig. 2). We note that at Mount St Helens the pressure–temperature trajectory crosses the amphibole-out boundary at around 120 MPa. Only melt inclusions from the plinian white pumice of 18 May 1980 lie entirely within the amphibole stability field, consistent with the absence of amphibole breakdown rims in these samples²⁰. All other samples show a mixture of pressures and temperatures within and without the amphibole stability field, consistent with variable extents of amphibole reaction rim development in these more slowly ascending, less explosive magmas²⁰.

Independent support for the observed temperature rise comes from thermometry of touching Fe–Ti oxides, which are present in all eruptions studied, as phenocrysts, groundmass microlites, and inclusions in other phenocrysts. At each volcano, the temperature spread recorded by the oxides (Fig. 2c, f) closely matches that from plagioclase–melt thermometry of melt inclusions (Fig. 2a, d). Oxide

Table 1 | Samples studied

Sample number	Description	Date erupted	Alias
Mount St Helens*			
SH10†	Dense juvenile material from phreatic explosion	10 April 1980	
USNM 115379-34	Cryptodome magma in lateral blast deposit	18 May 1980	34
SH80D	Cryptodome magma in lateral blast deposit	18 May 1980	
C85-310	Pale grey (microlite-bearing) pumice‡		
KC518PFB	Microlite-free white pumice from pyroclastic flow	18 May 1980	PLZ, KCPL, May
May25†	Airfall pumice	25 May 1980	
KC612PF	Pumice from pyroclastic flow	12 June 1980	Jun
KC722U	Pumice from pyroclastic flow	22 July 1980	Jul
KC807B	Pumice from pyroclastic flow	7 August 1980	Aug
SHKB23S	Dense pumice in levee of 10/80 pyroclastic flow	7 August 1980	
USNM 115418-60-2	Dome fragment	16 October 1980	
USNM 115418-60	Dome fragment	16 October 1980	60
USNM 115418-42	Pumice	16 October 1980	
USNM 115418-61	Dome	16 October 1980	
USNM 115427-1	Pumice	27 December 1980	
USNM 115427-4	Dome	27 December 1980	
USNM 115465	Dome	18 June 1981	
KC681	Dome	18 June 1981	
USNM 115773-18	Dome	19 March 1982	
USNM 115773-3	Pumice	19 March 1982	
Shiveluch 			
SHIV01/#1	Dome fragment	May 2001	
SHIV01/#2	Pumice	May 2001	
SHIV01/#3	Dome fragment	July 2001	
SHIV01/#4	Pumice	May 2001	
Shv202002	Dome fragment	July 2002	
Shv202003	Pumice	May 2001	

*USNM samples come from the Smithsonian Institution; where the same samples are designated with an alias in ref. 5 this is also given. All samples, except KC518PFB, contain microlite crystals in the groundmass.

†Source: Cascades Volcano Observatory.

‡Erupted during the initial basal pyroclastic flow (phase I) of the plinian eruption³². Likely remnant of cryptodome or conduit magmas.

§Sample location UTM 10562602E 5117871N.

||Samples described in ref. 31.

inclusions in phenocrysts provide the most reliable estimate of conditions at the onset of magma ascent from the deep-seated magma reservoir (that is, $\sim 870^\circ\text{C}$ at Mount St Helens and $\sim 850^\circ\text{C}$ at Shiveluch). Oxide phenocryst pairs and groundmass microlites extend the temperature spread to higher values. As oxides chemically re-equilibrate to changes in temperature on a timescale of several days²¹, the observed spread is consistent with partial re-equilibration of these oxides during decompression crystallization. This suggests that microlite-rich lava domes, which typically crystallize at shallow levels, will show higher equilibration temperatures than explosively erupted, microlite-poor magma of the same composition. These findings are in contrast to interpretations of spreads in oxide temperature data observed at other volcanoes, which include heating by new pulses of hotter, more-mafic magma just before eruption^{21,22}, or thermal stratification within the sub-volcanic reservoir²³.

The rise in temperature during decompression crystallization can also be calculated thermodynamically for any hydrous andesite magma, provided that the variation in crystallinity with pressure and temperature is known. We have performed calculations for a

generic H_2O -saturated silicic andesite (Fig. 3a), similar in composition to Mount St Helens and Mount Pinatubo magmas, ascending from 300 MPa with an initial (crystal-free) temperature of 880°C and latent heat release of 2.3°C per 1% crystallized. We consider both linear and nonlinear (polynomial) variations in crystal fraction (X) between solidus and liquidus. At each pressure, we solve the heat balance between crystallization and latent heat release to determine the equilibrium temperature, ignoring the heat of vapour exsolution or adiabatic expansion, which would make a small reduction to the calculated temperatures⁴. The calculated increases in temperature (Fig. 3b) and crystallinity (Fig. 3c) during decompression are similar for the linear and polynomial cases, and closely match the increases shown in Fig. 2. Amphibole breakdown starts at approximately 150 MPa (Fig. 3b).

Our results demonstrate that temperature increase caused by latent heat release is likely in any hydrous magma that decompresses sufficiently slowly to permit crystallisation, and a phenomenon that has important implications for the textures and physical properties of hydrous silicic magmas. For example, calculated low-pressure melt viscosities²⁴ are reduced by a factor of 5–10 compared to the

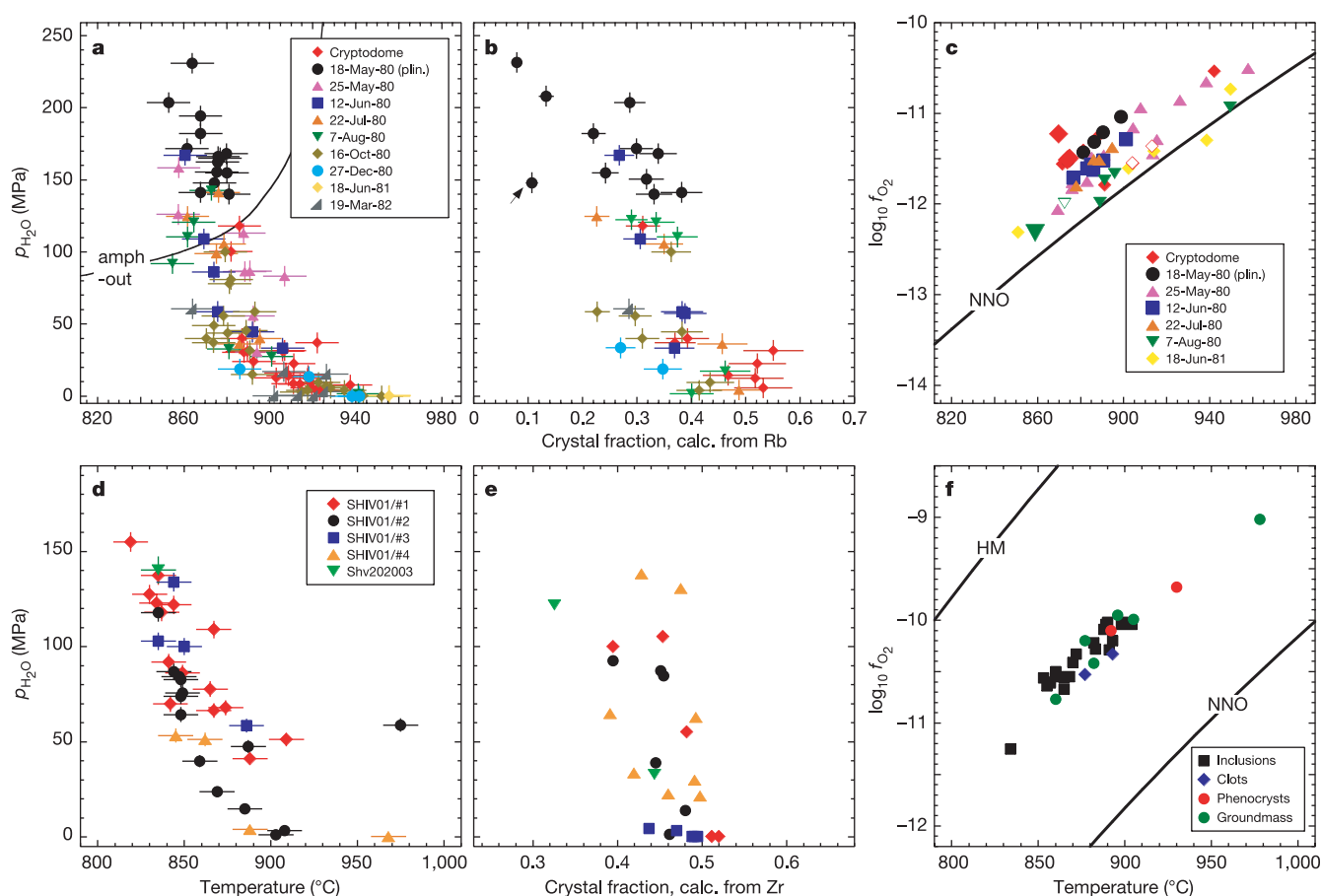


Figure 2 | Variation in magmatic variables. **a–c**, Beneath Mount St Helens; **d–f**, beneath Shiveluch. Data are taken from Supplementary Data Tables 1 and 2. Panels **a** and **d** show calculated $p_{\text{H}_2\text{O}}$ (from measured H_2O) and temperature (from plagioclase–melt thermometry) in plagioclase-hosted melt inclusions. Data are distinguished on the basis of eruption age (**a**) or sample number (**b**). In **a**, ‘Cryptodome’ includes pre-18 May 1980 juvenile material, the lateral blast deposit and early microlite-bearing plinian sample C85-310³². Error bars are ± 1 s.d. precision, as described in the text. The amphibole-out (‘amph-out’) curve in **a** is taken from ref. 20. Panels **b** and **e** show calculated variation in crystallinity as a function of $p_{\text{H}_2\text{O}}$. At Mount St Helens, crystallinity is calculated relative to the average whole-rock Rb content of 14 1980–86 dacites (30.8 ± 1.9 p.p.m.); at Shiveluch, we used the average whole-rock Zr of 112 ± 4 p.p.m. (ref. 31). Most of the scatter in **b**

and **e** results from using a constant whole-rock trace element content for the crystallinity calculations, when in reality there will be small variations between different magma batches. This is especially true for a single melt inclusion from Mount St Helens, shown with an arrow in **b**, which has anomalous concentrations of all trace elements and appears to represent an exotic melt component. Panels **c** and **f** show the calculated equilibration temperature and oxygen fugacity (f_{O_2}) for touching Fe–Ti oxide pairs. The nickel–nickel oxide (NNO) and magnetite–haematite (HM) oxygen buffers are shown for reference. In **c**, inclusions in silicate phenocrysts (large filled symbols), phenocryst oxides (filled symbols), and groundmass microlites (open symbols) are distinguished on the basis of eruption date; in **f**, they are distinguished on the basis of texture, including crystal-rich ‘clots’.

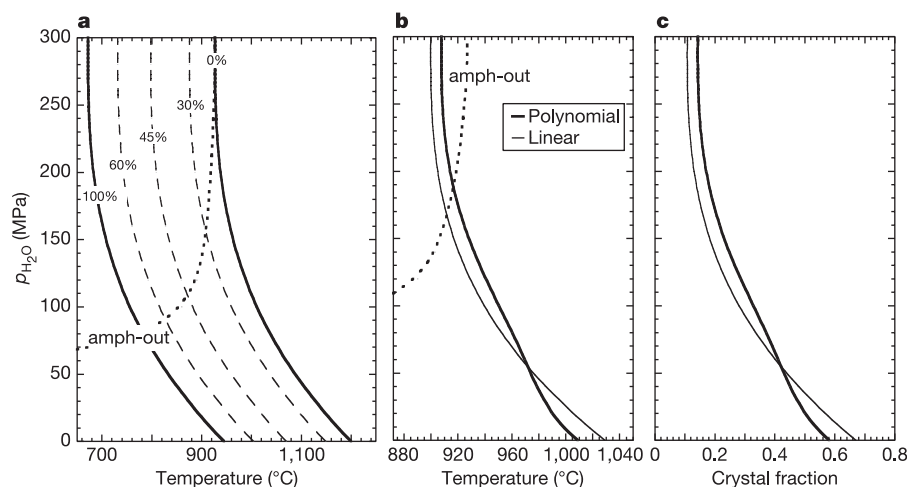


Figure 3 | Modelled variation in magmatic parameters during decompression of a generic H₂O-saturated silicic andesite. This material is similar in composition to Mount St Helens and Mount Pinatubo magmas. **a**, Phase diagram. Mount St Helens liquidus and haplogranite solidus parameterized from ref. 18. Crystal fraction (X , expressed as %) contours are based on a polynomial parameterization of 220 MPa H₂O-saturated experimental data on Mount Pinatubo¹⁶:

$$X = 2.757 \left(\frac{T_L - T}{T_L - T_S} \right)^3 - 3.922 \left(\frac{T_L - T}{T_L - T_S} \right)^2 + 2.165 \left(\frac{T_L - T}{T_L - T_S} \right) \quad (1)$$

isothermal case, which serves to offset the increase in viscosity that results from degassing and crystallization. Such changes in melt viscosity may affect the kinetics of crystal and bubble formation and the ease of gas migration out of magma conduits, thereby influencing both the rate and style of eruptive activity. The textural consequences of exothermic crystallization are equally far-reaching. For example, heating will lead to reverse-zoned rims on phenocrysts of plagioclase, Fe-Ti oxides and pyroxenes, which might otherwise be attributed to a heating event by hotter, more-mafic magma just before eruption^{21,22,25}.

Heating may also bear on oscillatory zoning in plagioclase phenocrysts²⁶ because of the opposing effects of decompression and heating on equilibrium liquidus plagioclase compositions, which serve, respectively, to decrease and increase anorthite (An) content. At Mount St Helens, we calculate that decompression decreases An by ~0.5 mol% per 1 MPa pressure drop, while latent heat release increases An by ~3 mol% for each per cent crystallized. The complex interplay between the two effects will result in a regime where crystals both grow in response to decompression and resorb due to heating by adjacent batches of crystallizing magma. For an individual crystal, the result will be a saw-tooth pattern characteristic of normal-oscillatory zoning²⁶, in which a short, normally-zoned step due to decompression crystallization is truncated by a resorption horizon and then overgrown by slightly higher-An plagioclase (Fig. 1a–c) as the sample becomes heated. A detailed quantitative evaluation of this proposal is beyond the scope of this paper. However, if our interpretation of oscillatory zoning is correct, then there is considerable scope for using the amplitude and frequency of oscillations in plagioclase to quantify magma ascent rates through the sub-volcanic system. Our results demonstrate the importance of considering latent heat release in the textural interpretation of magmatic phenocrysts and in models of sub-volcanic magma movement and eruption dynamics.

METHODS

Analytical methods. Melt inclusions were analysed for their major elements by a Cameca SX100 wavelength-dispersive electron-microprobe with a 2–4 nA, 15 kV defocused beam ($\geq 10 \mu\text{m}$ diameter) and reduced counting times for Na to

where T is temperature, T_L is the liquidus and T_S is the solidus at the pressure of interest. The amphibole-out (amph-out) curve is from ref. 20. **b**, Temperature based on linear variation of crystal fraction between solidus and liquidus (thin curve) and polynomial variation (equation (1), thick curve). **c**, Crystal fraction based on linear (thin curve) and polynomial (thick curve) variations. Calculations are based on an arbitrary initial temperature (crystal-free) at 300 MPa of 880 °C, which is slightly higher than that at either Mount St Helens or Shiveluch.

reduce alkali loss during irradiation²⁷. Oxide crystals were also analysed by electron-microprobe using a focused 10 nA, 20 kV beam. For oxide thermometry, only magnetite and ilmenite grains in direct contact with each other, or included in the same phenocryst, were analysed in order to sample the variability in magmatic temperatures within an individual sample. This is not possible with the conventional approach of using average analyses for thermometry. All analyses in our study meet the Mg-Mn equilibrium criterion of ref. 28. Oxide formulae were recalculated according to ref. 29; and temperature and oxygen fugacity were calculated using the method of ref. 30.

H₂O and light-element isotope (up to ⁴⁷Ti) contents of melt inclusions were analysed on a Cameca IMS4f ion-microprobe, using the method of ref. 5. Spot diameter at the sample surface was typically 8 μm for a 2 nA primary beam. A subset of the inclusions was analysed for heavy trace-element isotopes (⁴⁷Ti to ²³⁸U) by ion-microprobe, in a separate analytical session from H₂O and light elements to reduce magnet hysteresis. All trace elements were calibrated against NIST SRM610 glass; ³⁰Si, as determined by electron-microprobe analysis of the same spot, was used as the internal standard for all analyses. Analysis of secondary standards indicates that accuracy is within 10–15% relative for all elements analysed. Precision is always <5% relative, based on counting statistics. In this study, a requirement for acceptance of any melt inclusion analysis is that measurements of Ti from electron-probe, light-element ion-microprobe routine and heavy-element ion-microprobe routine agree to within $\pm 8\%$ relative, and that the analytical total lies between 98.2 and 101.5 wt%. Melt inclusions that showed chemical or textural evidence for syneruptive loss of H₂O (ref. 5), without concomitant crystallization, were excluded from this study. Plagioclase ingrowths into, or adjacent to, melt inclusions were analysed by energy-dispersive techniques on a Hitachi S-3500N scanning electron microscope calibrated against plagioclase standards of variable An content. Reproducibility is within ± 2.5 mol% An.

Melt inclusion analyses for Mount St Helens and Shiveluch are given in Supplementary Tables 1a and 1b, respectively; oxide analyses in Supplementary Tables 2a and 2b.

Received 15 March; accepted 17 July 2006.

- Cashman, K. V. in *The State of the Planet: Frontiers and Challenges in Geophysics* (eds Sparks, R. S. J. & Hawkesworth, C. J.) 109–124 (AGU Monograph 150, American Geophysical Union, Washington DC, 2004).
- Melnik, O. & Sparks, R. S. J. Non-linear dynamics of lava dome extrusion. *Nature* **402**, 37–41 (1999).
- Melnik, O. & Sparks, R. S. J. Controls on conduit magma flow dynamics during lava dome building eruptions. *J. Geophys. Res.* **110**, doi:10.1029/2004JB003183 (2005).

4. Mastin, L. G. & Ghiorsio, M. S. Adiabatic temperature changes of magma-gas mixtures during ascent and eruption. *Contrib. Mineral. Petrol.* **141**, 307–321 (2001).
5. Blundy, J. & Cashman, K. Rapid decompression-driven crystallisation recorded by melt inclusions from Mount St. Helens volcano. *Geology* **33**, 793–796 (2005).
6. Newman, S. & Lowenstern, J. B. VolatileCalc: a silicate melt–H₂O–CO₂ solution model written in Visual Basic for excel. *Comput. Geosci* **28**, 597–604 (2002).
7. Putirka, K. Igneous thermometers and barometers based on plagioclase + liquid equilibria: tests of some existing models and new calibrations. *Am. Mineral.* **90**, 336–346 (2005).
8. Costa, F., Scaillet, B. & Pichavant, M. Petrological and experimental constraints in the pre-eruption conditions of Holocene dacite from Volcán San Pedro (36°S, Chilean Andes) and the importance of sulphur in silicic subduction-related magmas. *J. Petrol.* **45**, 855–881 (2004).
9. Couch, S., Harford, C. L., Sparks, R. S. J. & Carroll, M. R. Experimental constraints on the conditions of formation of highly calcic plagioclase microlites at the Soufrière Hills volcano, Montserrat. *J. Petrol.* **44**, 1455–1475 (2003).
10. Hammer, J. E., Rutherford, M. J. & Hildreth, W. Magma storage prior to the 1912 eruption at Novarupta, Alaska. *Contrib. Mineral. Petrol.* **144**, 144–162 (2002).
11. Holtz, F., Sato, H., Lewis, J., Behrens, H. & Nakada, S. Experimental petrology of the 1991–1995 Unzen dacite, Japan. Part I: phase relations, phase compositions and pre-eruptive conditions. *J. Petrol.* **46**, 319–337 (2005).
12. Klimm, K., Holtz, F., Johannes, W. & King, P. L. Fractionation of metaluminous A-type granites: an experimental study of the Wangrah Suite, Lachlan Fold Belt, Australia. *Precamb. Res.* **124**, 327–341 (2003).
13. Larsen, J. F. Experimental study of plagioclase rim growth around anorthite seed crystals in rhyodacite melt. *Am. Mineral.* **90**, 417–427 (2005).
14. Martel, C. *et al.* Effects of fO₂ and H₂O on andesite phase relations between 2 and 4 kbar. *J. Geophys. Res.* **104**, 29453–29470 (1999).
15. Martel, C. & Schmidt, B. C. Decompression experiments as an insight into ascent rates of silicic magmas. *Contrib. Mineral. Petrol.* **144**, 397–415 (2003).
16. Scaillet, B. & Evans, B. W. The 15 June 1991 eruption of Mount Pinatubo. I. Phase equilibria and pre-eruption P–T–fO₂–fH₂O conditions of the dacite magma. *J. Petrol.* **40**, 381–411 (1999).
17. Scandone, R. & Malone, S. D. Magma supply, magma discharge and readjustment of the feeding system of Mount St. Helens during 1980. *J. Volcanol. Geotherm. Res.* **23**, 239–262 (1985).
18. Blundy, J. & Cashman, K. Ascent-driven crystallisation of dacite magmas at Mount St. Helens, 1980–1986. *Contrib. Mineral. Petrol.* **140**, 631–650 (2001).
19. Couch, S., Harford, C. L., Sparks, R. S. J. & Carroll, M. R. Mineral disequilibrium in lavas explained by convective self-mixing in open magma chambers. *Nature* **411**, 1037–1039 (2001).
20. Rutherford, M. J. & Hill, P. M. Magma ascent rates from amphibole breakdown: An experimental study applied to the 1980–1986 Mount St. Helens eruption. *J. Geophys. Res.* **98**, 19667–19686 (1993).
21. Venezky, D. Y. & Rutherford, M. J. Petrology and Fe–Ti oxide reequilibration of the 1991 Mount Unzen mixed magma. *J. Volcanol. Geotherm. Res.* **89**, 213–230 (1999).
22. Devine, J. D., Rutherford, M. J., Norton, G. E. & Young, S. R. Magma storage region processes inferred from geochemistry of Fe–Ti oxides in andesitic magma, Soufrière Hills Volcano, Montserrat, WI. *J. Petrol.* **44**, 1375–1400 (2003).
23. Hildreth, W. Gradients in silicic magma chambers: implications for lithospheric magmatism. *J. Geophys. Res.* **86**, 10153–10192 (1981).
24. Hess, K. U. & Dingwell, D. B. Viscosities of leucogranite melts: a non-Arrhenian model. *Am. Mineral.* **81**, 1297–1300 (1996).
25. Bachman, O. & Dungan, M. A. Temperature-induced Al-zoning in hornblendes of the Fish Canyon magma, Colorado. *Am. Mineral.* **87**, 1062–1076 (2002).
26. Ginibre, C., Kronz, A. & Wörner, G. High-resolution quantitative imaging of plagioclase composition using accumulated backscattered electron images: new constraints on oscillatory zoning. *Contrib. Mineral. Petrol.* **142**, 436–448 (2002).
27. Humphreys, M. C. S., Kearns, S. L. & Blundy, J. D. SIMS investigation of electron-beam damage to hydrous, rhyolitic glasses: Implications for melt inclusion analysis. *Am. Mineral.* **91**, 667–679 (2006).
28. Bacon, C. R. & Hirschmann, M. M. Mg/Mn partitioning as a test for equilibrium between coexisting Fe–Ti oxides. *Am. Mineral.* **73**, 57–61 (1988).
29. Spencer, K. J. & Lindsley, D. H. A solution model for coexisting iron-titanium oxides. *Am. Mineral.* **66**, 1189–1201 (1981).
30. Andersen, D. J. & Lindsley, D. H. Internally consistent solution models for Fe–Mg–Mn–Ti oxides: Fe–Ti oxides. *Am. Mineral.* **73**, 714–726 (1988).
31. Dirksen, O. V. *et al.* The 2001–2004 dome-forming eruption of Shiveluch Volcano, Kamchatka: observation, petrological investigation and numerical modelling. *J. Volcanol. Geotherm. Res.* **155**, 201–226 (2006).
32. Criswell, C. W. Chronology and pyroclastic stratigraphy of the May 18, 1980 eruption of Mount St. Helens, Washington. *J. Geophys. Res.* **92**, 10237–10266 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.B. was supported by an NERC Senior Research Fellowship and M.H. by an NERC Studentship. Ion-microprobe analysis benefited greatly from the efforts of the IMF staff at Edinburgh University and from standard materials supplied by T. Sisson, P. King and R. Brooker. W. Melson and D. Pyle provided some Mount St Helens samples.

Author Contributions The authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.B. (Jon.Blundy@bris.ac.uk).

Potential of cortical inhibition by visual deprivation

Arianna Maffei¹, Kiran Nataraj¹, Sacha B. Nelson¹ & Gina G. Turrigiano¹

The fine-tuning of circuits in sensory cortex requires sensory experience during an early critical period. Visual deprivation during the critical period has catastrophic effects on visual function, including loss of visual responsiveness to the deprived eye^{1–3}, reduced visual acuity⁴, and loss of tuning to many stimulus characteristics^{2,5}. These changes occur faster than the remodelling of thalamocortical axons⁶, but the intracortical plasticity mechanisms that underlie them are incompletely understood. Long-term depression of excitatory intracortical synapses has been proposed as a general candidate mechanism for the loss of cortical responsiveness after visual deprivation^{7,8}. Alternatively (or in addition), the decreased ability of the deprived eye to activate cortical neurons could be due to enhanced intracortical inhibition^{9,10}. Here we show that visual deprivation leaves excitatory connections in layer 4 (the primary input layer to cortex) unaffected, but markedly potentiates inhibitory feedback between fast-spiking basket cells (FS cells) and star pyramidal neurons (star pyramids). Further, a previously undescribed form of long-term potentiation of inhibition (LTPi) could be induced at synapses from FS cells to star pyramids, and was occluded by previous visual deprivation. These data suggest that potentiation of inhibition is a major cellular mechanism underlying the deprivation-induced degradation of visual function, and that this form of LTPi is important in fine-tuning cortical circuitry in response to visual experience.

In many animals, including rodents and primates, visual deprivation during a critical period induces a permanent loss of visual responsiveness and acuity referred to as amblyopia^{11,12}. In rodents, monocular deprivation during this critical period induces a rapid (within two days) loss of responsiveness to the deprived eye, followed more slowly (longer than five days) by increased responsiveness to the spared eye, indicating that these are separable processes that probably occur by distinct mechanisms³. Little is known about the synaptic changes that underlie the loss of responsiveness to the deprived eye. Here we used whole-cell recordings in rat visual cortical slices¹³ to selectively examine changes in excitatory and inhibitory microcircuitry within layer 4 after the suturing of one eyelid. In rodents, about two-thirds of the visual cortex is driven exclusively by the contralateral eye (monocular cortex), whereas a small binocular zone also receives a weak ipsilateral projection; most visual function is thus mediated solely by inputs from one eye. Visual deprivation (VD)-induced loss of cortical responsiveness and reduced visual acuity do not depend on competitive interactions between the two eyes^{4,11}, so the underlying mechanism can be studied in monocular cortex. Because in binocular cortex the synapses driven by the two eyes are intermingled, we restricted our analysis to monocular cortex, where we could probe the mechanisms underlying loss of responsiveness to the deprived eye by analysing connections that were unambiguously affected by the deprivation, and without

contamination from the slower potentiation of inputs from the non-deprived eye³.

To determine whether VD depresses excitatory synapses within layer 4, we examined monosynaptic connections between star pyramids, the major class of excitatory neuron within layer 4 of V1. VD between postnatal day 18 (P18) and P21 had no effect on excitatory postsynaptic current (EPSC) amplitude (Fig. 1a–c; control, $n = 73$; deprived, $n = 59$; $P = 0.57$), the probability of finding connected pairs (control, 10.7% of 739 tested; deprived, 11.4% of 553 tested), or on short-term plasticity (Fig. 1b; probed with trains of five spikes at 20 Hz, steady-state depression $42.3 \pm 4.5\%$ for control and $43.7 \pm 3.3\%$ for deprived; $P = 0.78$). Additionally, spike-timing-dependent long-term depression (LTD)¹⁴ was similar between hemispheres (Fig. 1d–f; $P = 0.43$). Taken together with our previous data showing no effect of VD on miniature EPSCs onto star pyramids at this developmental stage¹⁵, these data indicate that two days of VD during the critical period leaves excitatory transmission and plasticity onto star pyramids in layer 4 unaltered.

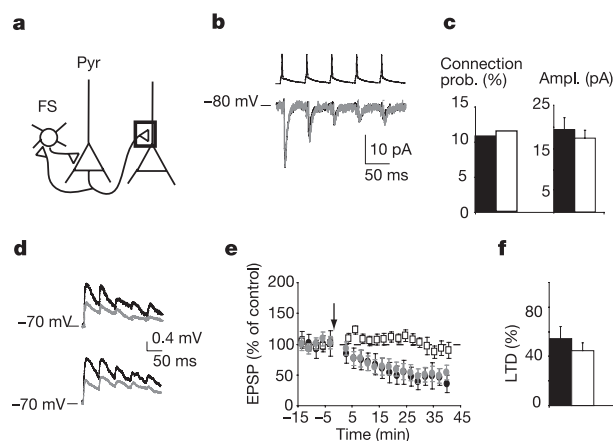


Figure 1 | VD between P18 and P21 has no effect on recurrent excitatory star pyramid connections. **a**, Diagram of excitatory star pyramid (Pyr) connections. **b**, Overlaid EPSCs from example control (black) and deprived (grey) pairs, showing presynaptic spike train (top) and postsynaptic current (bottom). **c**, Average EPSC amplitude (right) and connection probability (left) for control (filled bars) and deprived (open bars) pairs. **d**, Example LTD experiments (in current clamp) from control (top) and deprived (bottom) hemispheres (black, before pairing; grey, after pairing). **e**, Time course of EPSPs for control and deprived hemisphere pairs without LTD pairing protocol (white squares), and for control and deprived hemisphere pairs with LTD pairing protocol (black and grey circles, respectively). Arrow indicates the time of the induction protocol. **f**, Average magnitude of pairing-induced LTD for control (filled bar) and deprived (open bar) pairs. Where shown, error bars are s.e.m.

¹Department of Biology and Center for Behavioral Genomics, Brandeis University, Waltham, Massachusetts 02454, USA.

To analyse feedback inhibitory circuitry within layer 4, we obtained paired recordings between FS cells and star pyramids to analyse star pyramid to FS-cell excitation, and FS-cell to star-pyramid inhibition (Fig. 2). We focused on this microcircuit because FS cells are a major source of inhibition within layer 4, are important in regulating network excitability during the precritical period¹³ and have been implicated in the initiation of critical period plasticity¹⁶. FS cells receive much of their excitatory drive from star pyramids (Fig. 2a)¹⁷. VD induced a threefold increase in the amplitude of the star-pyramid to FS-cell EPSC (Fig. 2b, c; $P < 0.01$), accompanied by an increase in steady-state depression (control, to $42.9 \pm 6.6\%$ of initial value; deprived, to $69.2 \pm 6.7\%$; $P < 0.006$), and no significant change in connection probability (control, 57.4% of 47 tested; deprived, 53.5% of 41 tested). Thus, excitatory synapses within layer 4 are differentially regulated by VD according to their target: whereas connections from star pyramids to star pyramids remain constant, connections from star pyramids to FS cells are markedly increased in amplitude. This should serve to increase excitatory drive to FS cells and boost inhibition within layer 4.

We next determined the effect of VD on FS-cell to star-pyramid inhibitory synapses (Fig. 2d–f; the reversal potential was -49.6 ± 1.4 mV, close to the calculated chloride reversal potential, E_{Cl} , of -50.3 mV, and the holding potential was -80 mV, so currents were inward). There was a threefold increase in inhibitory postsynaptic current (IPSC) amplitude in the deprived hemisphere (Fig. 2e, f; control $n = 22$, deprived $n = 23$; $P < 0.01$). No significant differences were observed in connection probability (control, 50.0% of 50 tested; deprived, 61.9% of 42 tested) or short-term plasticity (steady-state depression: control, $71.5 \pm 4.9\%$; deprived, $63.5 \pm 5.4\%$; $P = 0.17$). The measured reversal potential was not different between control and deprived connections, when measured in whole-cell configuration (control, -49.6 ± 1.4 mV; deprived, -48.5 ± 1.8 mV), or with gramicidin-perforated patch recordings (Supplementary Fig. 1; control, -89.5 ± 0.7 mV; deprived, -88.5 ± 0.7 mV; $n = 3$). Last, there was no significant change in the strength of inhibition between FS cells (Supplementary Fig. 2; control, 21.9 ± 6.3 pA, $n = 4$; deprived, 26.1 ± 9.5 pA, $n = 3$).

Our data show that VD markedly and selectively potentiates both

components of the feedback inhibitory loop between star pyramids and FS cells, indicating that the balance between excitation and inhibition onto star pyramids has shifted to favour inhibition. Increased inhibition onto star pyramids should damp their activity; to test this we compared the spontaneous firing rates of star pyramids in slices derived from control and deprived cortex, as described previously¹³. We showed previously that VD during a precritical period just around eye opening (P14–P17) increases spontaneous activity of star pyramids¹³; in contrast, VD between P18 and P21 produced a more than tenfold reduction in spontaneous firing of star pyramids in the hemisphere driven by the deprived eye (Fig. 3; control, 3.34 ± 0.62 Hz, $n = 10$; deprived, 0.22 ± 0.07 Hz, $n = 10$; $P < 0.003$); a similar reduction was induced by VD between P21 and P24 (Fig. 3b). Although additional factors may contribute to this reduced excitability, a major cause is likely to be the potentiation of feedback inhibition.

Some inhibitory synapses can undergo LTPi^{18–20}, raising the possibility that VD increases the amplitude of FS-cell to star-pyramid connections by inducing LTP at this synapse. We used a current clamp protocol to induce LTPi reliably in the control hemisphere, by pairing FS-cell firing (trains of ten spikes at 50 Hz, repeated 20 times at 0.1 Hz) with subthreshold depolarization of the postsynaptic star pyramid (to between -60 and -55 mV; Fig. 4a, b, control). This reliably potentiated inhibitory postsynaptic potentials (IPSPs) to about 150% of baseline ($n = 9$; $P < 0.002$; all nine connections showed potentiation of more than 25%), with no changes in reversal potential (baseline, -47.6 ± 1.2 mV; induced, -48.7 ± 1.1 mV; $n = 9$; $P = 0.3$), and no change in paired-pulse depression (baseline, $35 \pm 2\%$; induced, $32 \pm 2\%$; $n = 9$; $P = 0.56$). Presynaptic firing alone (in the absence of postsynaptic depolarization) had no effect on IPSP amplitude (Fig. 4b, triangles); neither did postsynaptic depolarization alone ($89.7 \pm 7.7\%$ of control, $n = 3$; $P = 0.37$). In addition, various combinations of presynaptic and postsynaptic firing (presynaptic 10 ms before postsynaptic at 5 or 20 Hz, presynaptic 10 ms after postsynaptic at 20 Hz, presynaptic and postsynaptic together at 50 Hz), induced no change in synaptic strength (Fig. 4c, control, $n = 12$). Thus, LTPi at this synapse requires presynaptic spiking coupled with subthreshold postsynaptic depolarization but is prevented by coincident presynaptic and postsynaptic firing.

If LTPi underlies the VD-induced potentiation of inhibition they should share the same expression mechanism(s). To examine this more closely we performed an additional set of LTPi experiments ($n = 5$) in which we measured synaptic currents before and after LTPi (induced in current clamp as above), so that we could carefully analyse short-term plasticity and the coefficient of variation (CV). Comparing the CV of FS-cell to star-pyramid synapses in the control, deprived and LTPi cases revealed that both VD and LTPi significantly

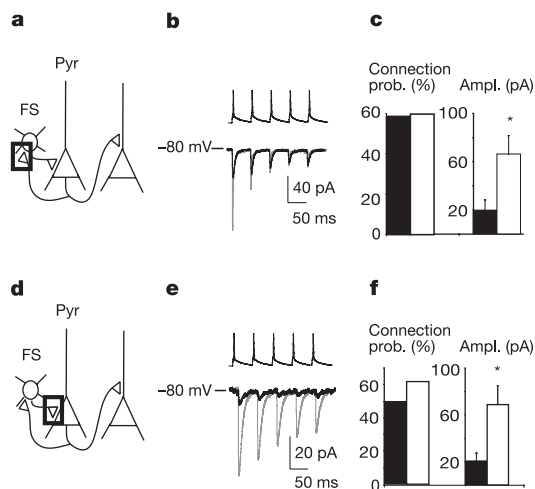


Figure 2 | VD potentiates feedback inhibition within layer 4. **a, b**, EPSCs from star pyramids (Pyr) onto FS neurons (**a**) were increased in amplitude by VD. **b**, Overlaid EPSCs from example control (black) and deprived (grey) pairs. **c**, Average connection probability and EPSC amplitude for control (filled bars) and deprived (open bars) pairs. **d, e**, IPSCs from FS cells onto star pyramids (**d**) were increased in amplitude by VD. **e**, Overlaid IPSCs from example control (black) and deprived (grey) pairs. **f**, Average connection probability (left) and IPSC amplitude (right) for control (filled bars) and deprived (open bars) pairs. Asterisk indicates statistical significance. Where shown, error bars are s.e.m.

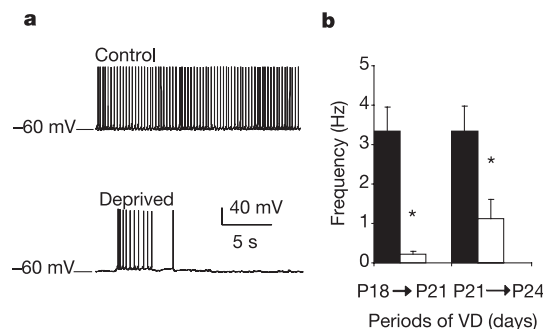


Figure 3 | VD suppresses spontaneous firing of star pyramids. **a**, Example recordings from star pyramids in slices derived from the control or deprived hemisphere, after VD between P18 and P21. **b**, Average firing rates for control (filled bars) and deprived (open bars) star pyramids after 2 days of VD during the ages indicated. Asterisk indicates statistical significance. Where shown, error bars are s.e.m.

reduced the CV (Fig. 4d, CV) but did not affect the paired-pulse ratio (Fig. 4d, PP ratio). To examine a possible postsynaptic contribution we performed non-stationary peak-scaled fluctuation analysis on the IPSCs (Supplementary Fig. 3). This revealed no difference in single-channel conductance between conditions (Fig. 4d, γ) but a significant increase in open channel number, as predicted if quantal content goes up; for both VD-induced potentiation and LTPi the increase in the number of open channels could fully account for the increase in current (Fig. 4e, f). Thus both forms of plasticity share the same expression characteristics.

A standard way of relating *in vitro* to *in vivo* plasticity phenomena is to ask whether the latter occludes the former: if VD increases

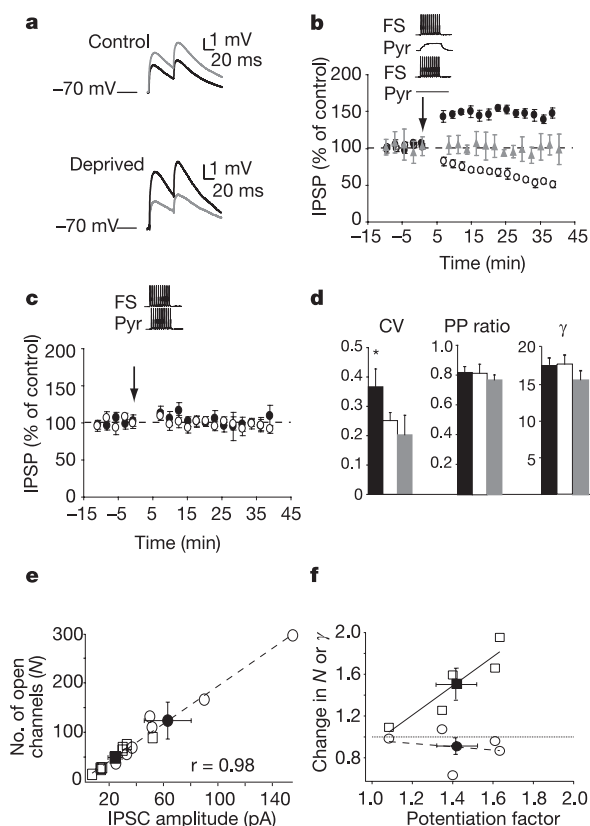


Figure 4 | Inhibitory LTP at FS-cell to star-pyramid synapses is occluded by previous VD. **a, b,** Pairing FS-cell firing with subthreshold star pyramid depolarization produced LTPi in the control hemisphere but LTDi in the deprived hemisphere. **a,** Top: example IPSPs from the control hemisphere before (black) and after (grey) pairing. Bottom: example IPSCs from the deprived hemisphere before pairing (black) and after pairing (grey). **b,** Average time course of IPSCs after pairing of FS cell firing with subthreshold star pyramid depolarization (inset, top) for control (filled circles) or deprived (open circles) connections. FS cell firing alone (inset, bottom) had no effect on IPSC amplitude (grey triangles). Arrows in **b** and **c** indicate time of the induction protocol. **c,** No plasticity was induced when star pyramid and FS cells were fired together. Filled circles, control pairs; open circles, deprived pairs. **d,** Bar plot showing CV, PP ratio and γ for control (black bars), deprived (white bars) and after LTPi induction (grey bars). Asterisk indicates statistical significance. **e,** Correlation between number of open channels (N) and IPSC amplitude for control (squares) and deprived (circles) pairs; each open symbol represents one pair; average values are represented by the filled symbols. The dashed line indicates the best linear fit ($r = 0.98$). **f,** Correlation between changes in N (open squares) or γ (open circles) and the potentiation factor (LTP amplitude divided by baseline amplitude) for each pair that underwent LTPi induction. Filled square, average N versus LTP factor; filled circle, average γ versus LTP factor. The solid line shows the best linear fit of N versus LTP factor ($r = 0.92$); the dashed line shows the best fit to γ versus LTP factor ($r = -0.2$). Where shown, error bars are s.e.m.

inhibitory transmission through LTPi, further induction of LTPi should be occluded. To test for such occlusion we compared LTPi in the control and deprived hemispheres. The same protocol that induced LTPi in the control hemisphere (Fig. 4a, b, control) did not induce LTP in the deprived hemisphere; rather, it induced a robust LTD (Fig. 4a, b, deprived, $n = 8$; $P < 0.01$) with no change in the reversal potential (control, -48.9 ± 0.9 mV; induced, -49.3 ± 0.8 mV; $n = 8$; $P = 0.7$) or paired-pulse depression. The induction protocols that were ineffective at inducing inhibitory plasticity in the control hemisphere were also ineffective in the deprived hemisphere (Fig. 4c, deprived, $n = 12$), indicating that the induction requirements might have been unaffected by VD. These data suggest that VD potentiates FS-cell to star-pyramid synapses through an LTPi-like process, which saturates LTPi and unmasks a competing LTD. Consistent with this was our observation that a second induction in the control hemisphere depressed synaptic transmission back to control levels, whereas in the deprived hemisphere it induced further LTD (Supplementary Fig. 4).

We have shown that a major effect of visual deprivation is a marked potentiation of feedback inhibition within layer 4 (see diagram, Supplementary Fig. 5). Our data support the idea that this occurs through an LTPi-like process that requires presynaptic firing coupled with subthreshold postsynaptic depolarization but is prevented by correlated presynaptic and postsynaptic firing. Because of these novel induction characteristics^{18–20}, LTPi at this synapse should only be induced when presynaptic and postsynaptic firing occur independently. Eye closure might produce just such an activity mismatch, because a decrease in the activity of star pyramids after eye closure should increase the probability that FS cells (which have a higher spontaneous firing rate^{21,22}) will fire in the absence of postsynaptic star pyramid firing. FS cells mediate inhibition that spreads mainly horizontally within a cortical layer²³, and are extensively electrically coupled²⁴. These properties allow them to inhibit a large population of excitatory neurons simultaneously. Increased feedback inhibition between FS cells and star pyramids should decrease the gain of incoming visual input by limiting the intracortical propagation and amplification of sensory information. As well as shaping cortical response properties, the inhibition of FS cells is important in initiating the critical period¹⁶, raising the possibility that inhibitory potentiation could be a necessary trigger for additional plasticity mechanisms such as those underlying the slower increase in responsiveness to the non-deprived eye³.

Some form of lateral, surround, or opponent inhibition is crucial for the function of many diverse microcircuits^{25–28}. In addition to contributing to the pathological changes in cortical function that follow sensory deprivation, the novel form of LTPi described here could serve as a general mechanism for sharpening lateral or opponent inhibition, because LTPi will strengthen inhibition onto neurons that are driven less effectively than the presynaptic interneuron by the same sensory stimuli. Finally, the ability of inhibitory plasticity to change sign as a function of previous history indicates that FS inhibition might be dynamically expanded or retracted by particular patterns of sensory input.

METHODS

General methods. Methods and solutions were as reported previously¹³. Monocular eyelid suture was performed late on P18 and animals were killed for recording early in P21; for some experiments (where indicated in the text) sutures were performed between P21 and P24. Coronal slices containing primary visual cortex were prepared from deprived and control hemispheres, and visualized patch clamp recordings were obtained from layer IV in the monocular region of V1. Neurons were filled with biocytin for post hoc reconstruction and were classified as described previously on the basis of morphology, firing properties, and synaptic properties¹³. Neurons were included if resting potentials were < -60 mV, input resistances were > 80 M Ω , series resistances were < 15 M Ω , and these parameters did not change more than 10% during the recording. Solutions were as described previously¹³.

Perforated patch recordings. Perforated patch recordings of star pyramids with

the use of gramicidin D were used to compare the GABA_A reversal potential between control and deprived hemispheres. Perforated patch internal solution had the following composition (in mM): KCl 130, EGTA 5, CaCl₂ 0.5, MgCl₂ 2, HEPES 10; pH made to 7.35 with KOH. Gramicidin D (50 µg ml⁻¹) was added to the internal solution before recordings and was sonicated for at least 30 s. The artificial cerebrospinal fluid for perforated patch recording contained DL-2-amino-5-phosphonopentanoic acid (50 µM) and 6,7-dinitroquinoxaline-2,3(1H,4H)-dione (20 µM). IPSCs were obtained by extracellular stimulation with a tungsten bipolar electrode (Harvard apparatus) positioned 40–100 µm from the recording site. The stimulation intensity was adjusted for each recording to the minimum needed to avoid failures. The equilibrium potential was obtained by measuring the voltage at the intersection between the holding-current and synaptic-current *I*–*V* curves as described in ref. 29 (Supplementary Fig. 1).

Plasticity induction protocols. Plasticity experiments were performed in current clamp to assess the role of presynaptic and postsynaptic firing in plasticity induction. A small bias current was injected into both presynaptic and postsynaptic neurons to keep their membrane potentials close to –70 mV between current injections. For LTD at recurrent star pyramid synapses, baseline synaptic transmission properties were assessed with trains of five presynaptic action potentials at 20 Hz, repeated at 0.05 Hz. During induction, presynaptic action potentials were paired, with a 10-ms delay, with postsynaptic action potentials (ten pairings at 0.1 Hz) as described¹⁴. For FS-cell to star-pyramid connections, baseline synaptic transmission properties were assessed with two presynaptic action potentials at 20 Hz, repeated at a frequency of 0.05 Hz. LTPi was successfully induced by pairing ten presynaptic action potentials at 50 Hz with postsynaptic depolarization to between –60 and –55 mV (20 pairings at 0.1 Hz). The protocols that failed to induce LTPi were as follows: 10-ms presynaptic before postsynaptic (control *n* = 3; deprived *n* = 3) or postsynaptic before presynaptic (control *n* = 3; deprived *n* = 3) pairing at 20 Hz (20 repetitions at 0.1 Hz); 10 ms presynaptic before postsynaptic pairing at 5 Hz for 30 s (ref. 19) (control *n* = 3, deprived *n* = 3); and presynaptic and postsynaptic firing together at 50 Hz (20 repetitions at 0.1 Hz; control *n* = 3; deprived *n* = 3). To perform noise analysis on IPSCs before and after LTPi induction, synaptic currents were recorded in voltage clamp; after a 10-min baseline, recordings were switched into current clamp for LTPi induction (pairing presynaptic firing at 50 Hz with postsynaptic depolarization to –60 mV as above); recordings were then switched back to voltage clamp to measure LTP of synaptic currents.

Non-stationary peak-scaled fluctuation analysis (NPSNA). NPSNA was performed using standard methods as described previously³⁰. In brief, for each unitary connection the average evoked IPSC was scaled to the peak of each individual IPSC and subtracted; the variance about the mean was then computed during the decay phase of the IPSCs (in 100-µs bins) and plotted against the mean current (Supplementary Fig. 3). The resulting relationship was well fitted with a parabolic equation from which could be extracted estimates of the single-channel conductance (γ) and the number of open channels (*N*). For LTPi experiments, this procedure was repeated for baseline currents and for post-induction currents.

Statistical tests. All data are expressed as mean $\pm$ s.e.m. for the number of pairs (or neurons) indicated. Statistical significance was determined by using two-tailed unpaired *t*-tests, except as follows. To determine the statistical significance of LTP and LTD within a condition, paired *t*-tests were performed on baseline amplitudes versus amplitudes 30 min after potentiation. The significance of changes in connection probability was tested with a χ^2 for contingency test.

Received 22 April; accepted 14 July 2006.

Published online 23 August 2006.

- Wiesel, T. N. & Hubel, D. H. Single cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* **26**, 1003–1017 (1963).
- Fagioli, M., Pizzorusso, T., Berardi, N., Domenici, L. & Maffei, L. Functional postnatal development of the rat primary visual cortex and the role of visual experience: dark rearing and monocular deprivation. *Vision Res.* **34**, 709–720 (1994).
- Frenkel, M. Y. & Bear, M. F. How monocular deprivation shifts ocular dominance in visual cortex of young mice. *Neuron* **44**, 917–923 (2004).
- Prusky, G. T., West, P. W. & Douglas, R. M. Experience-dependent plasticity of visual acuity in rats. *Eur. J. Neurosci.* **116**, 135–140 (2000).
- White, L. E., Coppola, D. M. & Fitzpatrick, D. The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature* **411**, 1049–1052 (2001).

- Antonini, A. & Stryker, M. P. Plasticity of geniculocortical afferents following brief or prolonged monocular occlusion in the cat. *J. Comp. Neurol.* **369**, 64–82 (1996).
- Rittenhouse, C. D., Shouval, H. Z., Paradiso, M. A. & Bear, M. F. Monocular deprivation induces homosynaptic long-term depression in visual cortex. *Nature* **397**, 347–350 (1999).
- Kirkwood, A., Rioult, M. C. & Bear, M. F. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* **381**, 526–528 (1996).
- Duffy, F. H., Burchfield, J. L. & Conway, J. L. Bicuculline reversal of deprivation amblyopia in the cat. *Nature* **260**, 256–257 (1976).
- Sillito, A. M., Kemp, J. A. & Blakemore, C. The role of GABAergic inhibition in the cortical effects of monocular deprivation. *Nature* **291**, 318–320 (1981).
- Rauschecker, J. P. Cortical map plasticity in animals and humans. *Prog. Brain Res.* **138**, 73–88 (2002).
- Barrett, B. T., Bradley, A. & McGraw, P. V. Understanding the neural basis of amblyopia. *Neuroscientist* **10**, 106–117 (2004).
- Maffei, A., Nelson, S. B. & Turrigiano, G. G. Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nature Neurosci.* **7**, 1353–1359 (2004).
- Egger, V., Feldmeyer, D. & Sakmann, B. Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in rat barrel cortex. *Nature Neurosci.* **2**, 1098–1105 (1999).
- Desai, N. S., Cudmore, R. H., Nelson, S. B. & Turrigiano, G. G. Critical periods for experience-dependent synaptic scaling in visual cortex. *Nature Neurosci.* **5**, 783–789 (2002).
- Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6**, 877–888 (2005).
- Thomson, A. M., Bannister, A. P., Mercer, A. & Morris, O. T. Target and temporal pattern selection at neocortical synapses. *Phil. Trans. R. Soc. Lond. B* **357**, 1781–1791 (2002).
- Gaiarsa, J. L., Caillard, O. & Ben-Ari, Y. Long-term plasticity at GABAergic and glycinergic synapses: mechanisms and functional significance. *Trends Neurosci.* **25**, 564–570 (2002).
- Woodin, M. A., Ganguly, K. & Poo, M. M. Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl[–] transporter activity. *Neuron* **39**, 807–820 (2003).
- Holmgren, C. D. & Zilberter, Y. Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells. *J. Neurosci.* **21**, 8270–8277 (2001).
- Simons, D. J. Response properties of vibrissa units in rat S1 somatosensory neocortex. *J. Neurophysiol.* **41**, 798–820 (1978).
- Contreras, D. & Palmer, L. Response to contrast of electrophysiologically defined cell classes in primary visual cortex. *J. Neurosci.* **23**, 6936–6945 (2003).
- Buzas, P., Eysel, U. T., Adorjan, P. & Kisvarday, Z. F. Axonal topography of cortical basket cells in relation to orientation, direction and ocular dominance maps. *J. Comp. Neurol.* **473**, 259–285 (2001).
- Gibson, J. R., Bierlein, M. & Connors, B. W. A network of electrically coupled inhibitory neurons in neocortex. *Nature* **402**, 75–79 (1999).
- Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbits' retina. *J. Physiol.* **178**, 477–504 (1965).
- Schoppa, N. E. & Urban, N. N. Dendritic processing within olfactory bulb circuits. *Trends Neurosci.* **26**, 501–506 (2003).
- Kayser, A. S. & Miller, K. D. Opponent inhibition: a developmental model of layer 4 of the neocortical circuit. *Neuron* **33**, 131–142 (2002).
- Hirsch, J. A. et al. Functionally distinct inhibitory neurons at the first stage of visual cortical processing. *Nature Neurosci.* **6**, 1300–1308 (2003).
- Jin, X., Huguenard, J. R. & Prince, D. A. Impaired Cl[–] extrusion in layer V pyramidal neurons of chronically injured epileptogenic neocortex. *J. Neurophysiol.* **93**, 2117–2126 (2005).
- Kilman, V., van Rossum, M. C. & Turrigiano, G. G. Activity deprivation reduces miniature IPSC amplitude by decreasing the number of postsynaptic GABA_A receptors clustered at neocortical synapses. *J. Neurosci.* **22**, 1328–1337 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Pavlyuk for help with histology, and A. Fontanini for help with software and for discussions. This study was supported by the National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.G.T. (turrigiano@brandeis.edu).

Experience-dependent representation of visual categories in parietal cortex

David J. Freedman¹ & John A. Assad¹

Categorization is a process by which the brain assigns meaning to sensory stimuli. Through experience, we learn to group stimuli into categories, such as 'chair', 'table' and 'vehicle', which are critical for rapidly and appropriately selecting behavioural responses^{1,2}. Although much is known about the neural representation of simple visual stimulus features (for example, orientation, direction and colour), relatively little is known about how the brain learns and encodes the meaning of stimuli. We trained monkeys to classify 360° of visual motion directions into two discrete categories, and compared neuronal activity in the lateral intraparietal (LIP) and middle temporal (MT) areas, two interconnected brain regions³ known to be involved in visual motion processing^{4–6}. Here we show that neurons in LIP—an area known to be centrally involved in visuo-spatial attention^{7–9}, motor planning^{10–13} and decision-making^{14–16}—robustly reflect the category of motion direction as a result of learning. The activity of LIP neurons encoded directions of motion according to their category membership, and that encoding shifted after the monkeys were retrained to group the same stimuli into two new categories. In contrast, neurons in area MT were strongly direction selective but carried little, if any, explicit category information. This indicates that LIP might be an important nexus for the transformation of visual direction selectivity to more abstract representations that encode the behavioural relevance, or meaning, of stimuli.

Monkeys were trained to group 12 directions of motion into two categories that were separated by a learned 'category boundary' (Fig. 1a, black dotted line). Monkeys performed a delayed-match-to-category (DMC) task (Fig. 1b) in which they viewed a sample stimulus (650 ms) followed by a delay (1,000 ms) and a test stimulus (650 ms). On a given trial, the sample and test could each be any one of the 12 directions of motion (see Methods). To receive a reward, the monkeys had to release a lever if the test was in the same category as the sample. If the test was a non-match, another delay (150–250 ms) occurred; this was followed by an additional test (650 ms), which was always a match to the sample (and required a lever release). By using this task design, lever releases signalled 'match' and were not directly linked to either category.

After training, the monkeys correctly categorized sample stimuli that were 75° or 45° from the category boundary with greater than 90% average accuracy, and performed at more than 70% correct for stimuli closest to (15°) the boundary (Fig. 1c).

We recorded from a total of 156 LIP neurons from two monkeys (monkey S, $n = 92$; monkey H, $n = 64$) during DMC task performance. A striking number of these neurons were category selective: their activity reliably grouped the 12 motion directions according to their category membership. Figure 2 shows the activity of three category-selective LIP neurons. The 12 traces correspond to the 12 motion directions used as samples, and are coloured red or blue according to their category membership. The pale red and blue traces indicate the four directions closest to (15°) the category boundary.

The neuron in Fig. 2a responded more strongly, during the delay, to directions in category 2. The neuron in Fig. 2b preferred sample directions in category 2 during the sample, delay and test, whereas the neuron in Fig. 2c preferred directions in category 1 during the sample, delay and test. Note that each of these neurons showed sharper (that is, binary-like) category selectivity during the delay than during the sample.

For quantitative analyses of neuronal activity, we focused on three time epochs, the 'sample', 'delay' and 'test'. The sample epoch (675 ms duration) began 75 ms after sample onset and ended 100 ms after sample offset. The delay epoch (800 ms duration) began 500 ms after the beginning of the delay (to exclude responses related to sample offset) and included the first 300 ms of the test epoch (because many LIP neurons carried information about the sample stimulus even during the test epoch; see below). Selectivity for the test stimulus

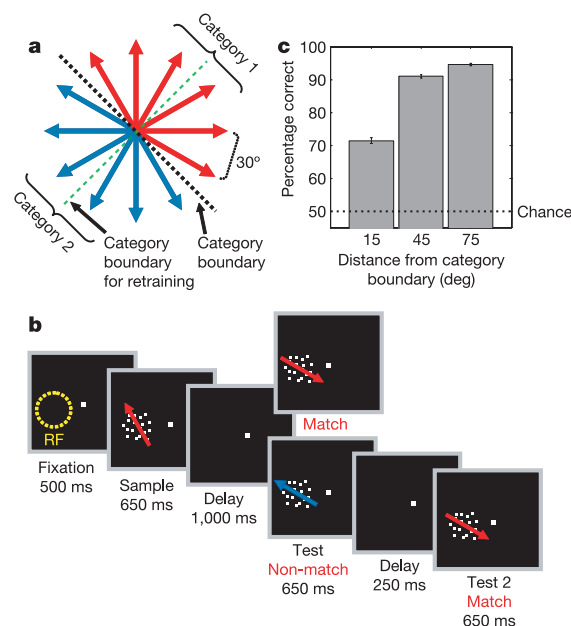


Figure 1 | Behavioural task. **a**, Monkeys grouped 12 motion directions into two categories (the red and blue arrows) separated by a 'category boundary' (black dotted line). The green dotted line is the boundary used for retraining with the new categories. **b**, Delayed match-to-category (DMC) task. A sample stimulus was followed by a delay and test. If the sample and test were in the same category, monkeys were required to release a lever before the test disappeared. If the test was a non-match, there was a second delay followed by a match (which required a lever release). **c**, Monkeys' average DMC task performance across all recording sessions was greater than chance (50%) for sample stimuli that were close to (15°) and farther from (45° or 75°) the boundary. Error bars are s.e.m.

¹Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.

and match/non-match effects were analysed over an interval of 275 ms beginning 75 ms after test stimulus onset (Supplementary Information).

A majority of LIP neurons (122 of 156; 78%) showed activity that differed across the 12 motion directions during the sample ($n = 115$) and/or delay ($n = 61$; one-way analysis of variance (ANOVA) across 12 directions, $P < 0.01$). However, this direction selectivity was closely related to the distinction between categories. To evaluate whether individual neurons responded more similarly to directions within than between categories, we computed two parameters: a within-category difference (WCD) and a between-category difference (BCD) in average firing rates to the 12 sample directions (Methods, and Supplementary Information). Across the entire LIP population ($n = 156$), responses to directions in the same category were more similar than to directions in different categories. This was evident during both the sample (Fig. 3a: WCD, 6.22 Hz; BCD, 8.72 Hz; paired t -test, $P \approx 10^{-9}$) and delay (Fig. 3b: WCD, 2.62 Hz; BCD, 4.42 Hz; $P \approx 10^{-11}$).

We also recorded from 67 middle temporal (MT) neurons (monkey S, $n = 40$; monkey H, $n = 27$) during DMC task performance. During the sample, nearly all (66 of 67; 99%) MT neurons distinguished between the 12 directions (one-way ANOVA, $P < 0.01$). However, MT responses did not systematically discriminate between categories (Supplementary Fig. 1). Across the entire MT population, WCD and BCD values were not significantly different during the sample (WCD, 21.49 Hz; BCD, 22.66 Hz; paired t -test, $P = 0.61$; Supplementary Fig. 2a). During the delay, no MT

neurons (0 of 67) were selective for the direction of the previously presented sample, and WCD and BCD did not differ from one another (WCD, 2.92 Hz; BCD, 2.93 Hz; paired t -test, $P = 0.93$; Supplementary Fig. 2b).

To measure the strength of neuronal category selectivity, we constructed a category-tuning index by taking the difference between BCD and WCD and dividing by their sum. Category-index values can vary from -1.0 to 1.0 , where positive values indicate larger differences for directions in different categories and negative values larger differences within each category (Methods and Supplementary Information). The distributions of category-tuning indices across the entire LIP population ($n = 156$) are shown in Fig. 3c, d. During both epochs, the mean category indices were shifted towards positive values (sample: mean 0.125, t -test, $P \approx 10^{-11}$; delay: mean 0.180, $P \approx 10^{-15}$), with stronger category tuning during the delay than sample (paired t -test, $P = 0.019$). The positive shift of LIP category indices indicates that the distribution of preferred directions became highly non-uniform as a result of training in the categorization task (see below and Methods). In addition, we did not detect an obvious relationship between LIP activity and reward probability (Supplementary Information).

In contrast, category tuning was not observed across the MT population ($n = 67$ neurons; Supplementary Fig. 1). Mean category-

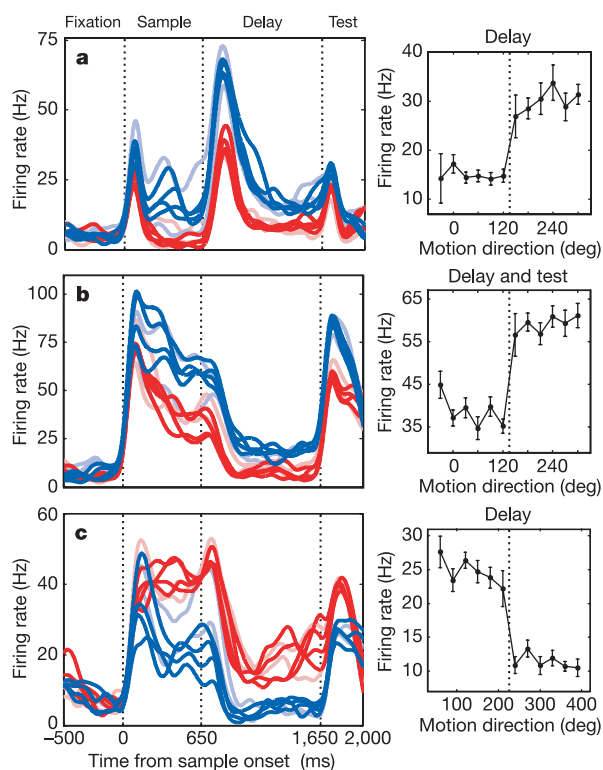


Figure 2 | Examples of three category-selective LIP neurons. Average activity to the 12 sample directions for three LIP neurons is shown. The red and blue traces correspond to directions in the two categories (red, category 1; blue, category 2), and pale traces indicate the directions closest to (15°) the boundary. The three vertical dotted lines indicate (from left to right) the timing of sample onset, the sample offset and the test-stimulus onset. The neurons in **a** and **b** were recorded with the original category boundary. The neuron in **c** was recorded after the monkey had been retrained on the new categories. The plots at the right of each peri-stimulus time histogram (PSTH) show activity (means $\pm$ s.e.m.) for the 12 directions during the delay (**a**), delay and test (**b**) and delay (**c**).

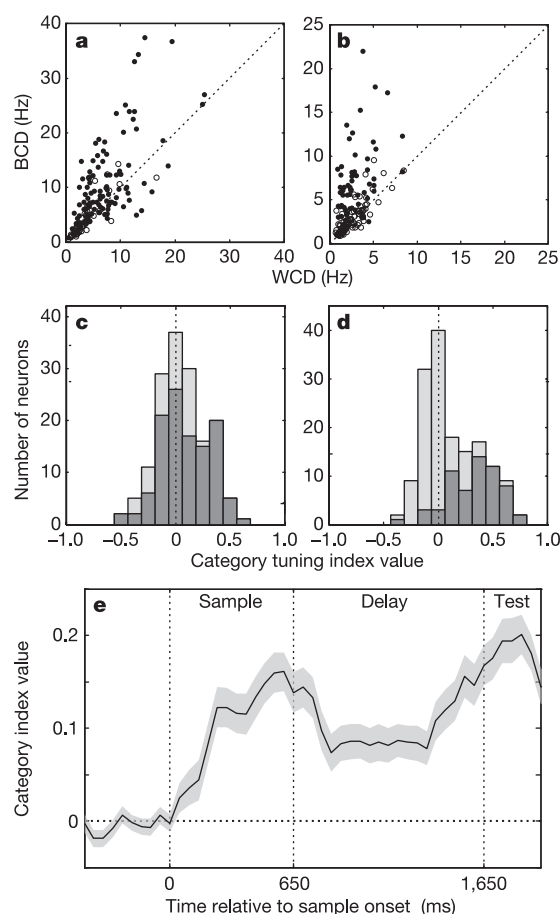


Figure 3 | Category effects across the LIP population. **a, b**, For each neuron, BCD and WCD values are shown for the sample (**a**) and delay (**b**). The filled and open circles indicate direction-selective and non-direction-selective neurons, respectively. **c, d**, A category index was computed from the BCD and WCD values. Positive index values indicate greater selectivity between categories and/or more similar activity within categories. The light grey bars show index values across all neurons ($n = 156$). The dark grey bars show index values for direction-selective neurons during sample (**c**) and delay (**d**). **e**, The time course of average category indices across 122 direction-selective neurons (during sample and/or delay).

tuning indices during both the sample (mean index 0.015) and delay (mean index 0.008) were not significantly different from zero (t -tests; sample, $P = 0.73$; delay, $P = 0.61$; Supplementary Fig. 2c, d).

To assess the time course of LIP category tuning in more detail, we computed a 'sliding' category-tuning index (window width 100 ms, step size 50 ms). Figure 3e shows the time course for 122 neurons that were direction-selective during the sample and/or delay. Category effects were evident within 100 ms of sample onset and persisted throughout the sample. After sample offset, category tuning waned somewhat but became progressively stronger across the delay, reaching peak values during the late delay and early test, when the monkey presumably prepared to compare the test category with that of the previously presented sample. It is notable that information about the category of the previously presented sample stimulus was strongest during the early test epoch, although the sample was presented 1 s earlier and the monkey was currently viewing the test stimulus.

Stronger category-tuning indices during the delay were apparently due to tighter clustering of responses to directions within each category and larger differences between categories. To confirm this, for each neuron we computed two separate one-way ANOVAs ($P < 0.01$, with Bonferroni correction) that compared responses to the six directions in each category. For the 115 LIP neurons that were direction selective across all 12 directions during the sample, the majority (92 of 115; 80%) also distinguished between the six directions within one or both of the two categories. However, for the 61 direction-selective neurons during the delay, a significantly smaller proportion (26 of 61, or 43%) differentiated between the six directions within either category (χ^2 test, sample versus delay, $P = 0.02$). Thus the responses tended to become more 'binary'

during the delay, reflecting category membership. This trend is evident in the example neurons in Fig. 2a–c. In contrast, nearly all MT neurons (66 of 67; 99%) were selective between the six directions in either category during the sample epoch.

To ensure that LIP category effects were due to learning the DMC task, we retrained both monkeys to group the same 12 directions into two new categories separated by a category boundary perpendicular to the original boundary (Fig. 1a, green dotted line). LIP selectivity shifted markedly with retraining. After retraining, neurons reflected the new categories and not the old (now irrelevant) categories. To quantify this effect, we determined which of six possible category boundaries (which divided the 12 directions into two equal groups) resulted in the greatest difference between average neuronal activity among the six directions on each side of the boundary. Among neurons recorded with the original boundary (92 neurons from both monkeys), sample and delay activity for most neurons was best classified by the actual category boundary that the monkeys were using and not the other five 'irrelevant' boundaries (Fig. 4a, b). After retraining the monkeys on the new categories, neuronal activity (across 64 neurons tested with the new category boundary) no longer reflected the old category boundary but was best divided by the new, now relevant, boundary (Fig. 4c, d).

Together these results indicate that training monkeys to perform a motion-categorization task causes neurons in LIP to strongly and robustly reflect the category membership of visual motion direction. LIP neurons responded more similarly to motion directions of the same category even when those directions were visually dissimilar, and they discriminated sharply between visually similar directions of different categories. In contrast, neurons in area MT, an important stage of visual motion processing⁴ that provides input to LIP³, were highly direction selective but did not group directions according to their category membership.

After retraining the monkeys to group the same stimuli into two new categories, LIP selectivity shifted markedly to encode the motion directions according to the newly learned categories. This demonstrates a profound learning-based plasticity of visual representations in LIP, beyond that typically seen in striate or extrastriate visual cortex¹⁷, and indicates that LIP is probably important in encoding the behavioural relevance, or meaning, of visual-motion stimuli. The exact nature of the role of LIP during learning, and whether changes in the directional representations of LIP are stable or vary dynamically with the demands of the task, remain to be determined.

Whereas recent studies found categorical representations in the prefrontal cortex^{18,19}, a frontal lobe area involved in more 'executive' functions²⁰, it has been unclear whether neurons in brain areas considered to be more involved in sensory processing could encode category information. For example, studies of shape categorization in inferior temporal cortex found enhanced selectivity for task-relevant features or shapes^{21,22} as a result of learning, but did not show more explicit signals about category membership^{23,24}. Because area LIP is known to be involved in both sensory²⁵ and cognitive functions^{26–28}, it is well positioned for a function in transforming sensory information to more abstract, and meaningful, representations of visual stimuli.

METHODS

Physiological techniques. Two male rhesus monkeys (*Macaca mulatta*, weighing about 14 kg) were implanted with a head post, scleral search coil and recording chamber. Recording chambers were implanted in accordance with coordinates (approximate centres at P3, L10) determined by magnetic resonance imaging, and allowed access to both the intraparietal sulcus (IPS) and superior temporal sulcus by means of a dorsal approach. All surgical and experimental procedures followed Harvard Medical School and National Institutes of Health guidelines.

During LIP recordings, electrode penetrations sequentially encountered both the medial and lateral banks of the IPS. Most IPS neurons were tested with a memory-saccade task and a passive viewing flash-mapping task to generate detailed spatial maps of neuronal response fields (RF). Neurons were considered

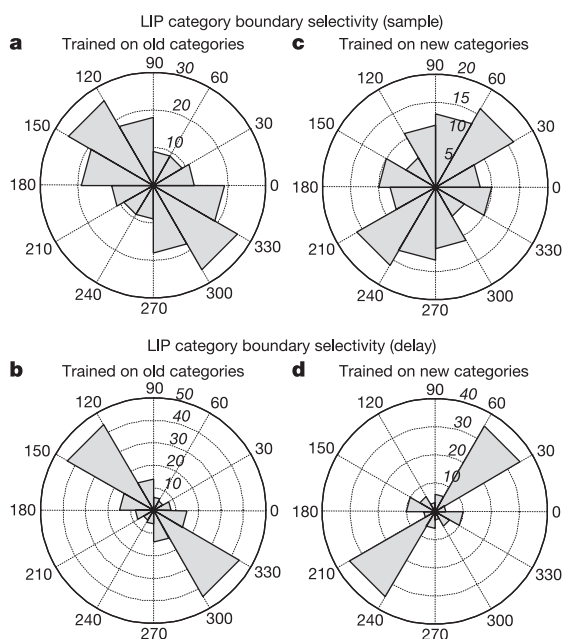


Figure 4 | LIP category selectivity followed retraining. After the recording of 92 neurons with the original category boundary, monkeys were retrained to categorize the same stimuli with a boundary perpendicular to the original one; we then recorded an additional 64 neurons. For each neuron we determined which of six possible boundaries (that is, the 'new' and 'old' boundaries, plus the four boundaries that we did not use) gave the largest difference in average activity among directions on either side of the boundary. **a, b**, Polar distribution of the best boundary for the sample (**a**) and delay (**b**) activity of all neurons with monkeys trained on the original boundary. The number of neurons that preferred each boundary (measured on the concentric scale with numbers in *italics*) corresponds to the radius of either of the two opposite 'bowtie' segments (but the two should not be added). **c, d**, Distribution of the best boundary for the activity of all neurons after retraining on the new categories.

to be in LIP if they showed spatially selective delay activity during the memory saccade task or were located between such neurons in that electrode penetration. LIP neurons were not prescreened for direction selectivity. Area MT neurons were distinguished by direction-selective responses to moving spots and bars, and RF sizes that were roughly proportional to their eccentricity⁴.

DMC task. Monkeys were trained to indicate whether a test stimulus was in the same category as a previously presented sample stimulus. The monkeys could not predict whether a given trial would require a release to the first test stimulus. Monkeys' average reaction times on correct match trials were 349 ms (monkey S) and 368 ms (monkey H).

Stimuli were circular patches (9.0° in diameter) of high-contrast square dots that moved in 1 of 12 evenly spaced directions (30° apart) with 100% motion coherence and at a speed of 12.0° s⁻¹. Stimuli were always centred in the RF of the neuron under study. Trials began with the onset of a 0.25° spot, which monkeys were required to fixate within ±1.5° for the duration of the trial. During recordings from monkey H we excluded the four test stimuli closest to (15°) the category boundary where the monkey made the greatest number of errors.

Data analysis. All analyses (except error trial analyses) were conducted across correct trials. The pattern of behavioural and neuronal results was similar, and all main effects were observed in both monkeys. The precise timing of analysis windows was not critical; similar results were obtained with a variety of window widths and starting points.

The strength of neuronal category selectivity was estimated by a category-tuning index. It was determined for each neuron by computing two values: the average WCD in firing rates between pairs of directions in the same category, and the average BCD in firing rates between pairs of directions in different categories (Supplementary Information). The index was computed with the formula $(BCD - WCD)/(BCD + WCD)$ and could vary from -1.0 to 1.0. The index was computed according to the currently relevant category boundary, allowing the data sets with each category boundary to be combined and analysed together.

Received 5 May; accepted 13 July 2006.

Published online 27 August 2006.

- Ashby, F. G. & Maddox, W. T. Human category learning. *Annu. Rev. Psychol.* **56**, 149–178 (2005).
- Barsalou, L. W. *Cognitive Psychology: An Overview for Cognitive Scientists* (Erlbaum, Hillsdale, New Jersey, 1992).
- Lewis, J. W. & Van Essen, D. C. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137 (2000).
- Born, R. T. & Bradley, D. C. Structure and function of visual area MT. *Annu. Rev. Neurosci.* **28**, 157–189 (2005).
- Eskandar, E. N. & Assad, J. A. Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nature Neurosci.* **2**, 88–93 (1999).
- Williams, Z. M., Elfar, J. C., Eskandar, E. N., Toth, L. J. & Assad, J. A. Parietal activity and the perceived direction of ambiguous apparent motion. *Nature Neurosci.* **6**, 616–623 (2003).
- Colby, C. L., Duhamel, J. R. & Goldberg, M. E. Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiol.* **76**, 2841–2852 (1996).
- Colby, C. L. & Goldberg, M. E. Space and attention in parietal cortex. *Annu. Rev. Neurosci.* **22**, 319–349 (1999).
- Bisley, J. W. & Goldberg, M. E. Neuronal activity in the lateral intraparietal area and spatial attention. *Science* **299**, 81–86 (2003).
- Snyder, L. H., Batista, A. P. & Andersen, R. A. Coding of intention in the posterior parietal cortex. *Nature* **386**, 167–170 (1997).
- Batista, A. P., Buneo, C. A., Snyder, L. H. & Andersen, R. A. Reach plans in eye-centered coordinates. *Science* **285**, 257–260 (1999).
- Buneo, C. A., Jarvis, M. R., Batista, A. P. & Andersen, R. A. Direct visuomotor transformations for reaching. *Nature* **416**, 632–636 (2002).
- Andersen, R. A. & Buneo, C. A. Intentional maps in posterior parietal cortex. *Annu. Rev. Neurosci.* **25**, 189–220 (2002).
- Platt, M. L. & Glimcher, P. W. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238 (1999).
- Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
- Huk, A. C. & Shadlen, M. N. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* **25**, 10420–10436 (2005).
- Ghose, G. M. Learning in mammalian sensory cortex. *Curr. Opin. Neurobiol.* **14**, 513–518 (2004).
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *J. Neurophysiol.* **88**, 914–928 (2002).
- Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
- Sigala, N. & Logothetis, N. K. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415**, 318–320 (2002).
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb. Cortex* (in the press).
- Vogels, R. Categorization of complex visual images by rhesus monkeys. *Eur. J. Neurosci.* **11**, 1223–1238 (1999).
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* **23**, 5235–5246 (2003).
- Bisley, J. W., Krisna, B. S. & Goldberg, M. E. A rapid and precise on-response in posterior parietal cortex. *J. Neurosci.* **24**, 1833–1838 (2004).
- Toth, L. J. & Assad, J. A. Dynamic coding of behaviorally relevant stimuli in parietal cortex. *Nature* **415**, 165–168 (2002).
- Nieder, A. & Miller, E. K. A parieto-frontal network for visual numerical information in the monkey. *Proc. Natl Acad. Sci. USA* **101**, 7457–7462 (2004).
- Stoet, G. & Snyder, L. H. Single neurons in posterior parietal cortex (PPC) of monkeys encode cognitive set. *Neuron* **42**, 1003–1012 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Ditterich, A. Fanini, V. Ferrera, J. Fitzgerald, C. Freedman, T. Herrington, M. Histed, G. Maimon, E. Miller, C. Pack, C. Padoa-Schioppa, A. Seitz and J. Wallis for comments, help and discussions, and K. Irwin, T. Lafratta and J. LeBlanc for technical assistance. This work was supported by the National Eye Institute (NEI) and the McKnight Endowment Fund for Neuroscience, and a Kirschstein postdoctoral National Research Service Award from the NEI to D.J.F.

Author Contributions D.J.F. performed all aspects of this study including the experimental design, data collection and analysis, and writing the manuscript. J.A.A. assisted in experimental design, data analysis and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J.F. (davidfreedman@alum.mit.edu).

Interference among deleterious mutations favours sex and recombination in finite populations

Peter D. Keightley¹ & Sarah P. Otto²

Sex and recombination are widespread, but explaining these phenomena has been one of the most difficult problems in evolutionary biology. Recombination is advantageous when different individuals in a population carry different advantageous alleles^{1,2}. By bringing together advantageous alleles onto the same chromosome, recombination speeds up the process of adaptation^{1,3–5} and opposes the fixation of harmful mutations by means of Muller's ratchet^{4,5}. Nevertheless, adaptive substitutions favour sex and recombination only if the rate of adaptive mutation is high^{1,6}, and Muller's ratchet operates only in small or asexual populations⁷. Here, by tracking the fate of modifier alleles that alter the frequency of sex and recombination, we show that background selection against deleterious mutant alleles provides a stochastic advantage to sex and recombination that increases with population size. The advantage arises because, with low levels of recombination, selection at other loci severely reduces the effective population size and genetic variance in fitness at a focal locus⁸ (the Hill–Robertson effect), making a population less able to respond to selection and to rid itself of deleterious mutations. Sex and recombination reveal the hidden genetic variance in fitness by combining chromosomes of intermediate fitness to create chromosomes that are relatively free of (or are loaded with) deleterious mutations. This increase in genetic variance within finite populations improves the response to selection and generates a substantial advantage to sex and recombination that is fairly insensitive to the form of epistatic interactions between deleterious alleles. The mechanism supported by our results offers a robust and broadly applicable explanation for the evolutionary advantage of recombination and can explain the spread of costly sex.

Sex and recombination break apart co-adapted gene combinations, thereby reducing fitness. Yet sexual reproduction is extremely widespread, and most eukaryotes have at least one chiasma per chromosome per meiosis⁹. One of the most promising evolutionary explanations for sex and recombination is that they break down associations between deleterious and beneficial alleles at different loci (negative disequilibria). By bringing together favourable alleles from different chromosomes, sex and recombination increase the additive genetic variance for fitness, and, by Fisher's fundamental theorem of natural selection, can increase the rate of adaptation^{1,3,4,10,11}. Nevertheless, current evolutionary models have difficulties in explaining widespread sex and recombination by this mechanism. Directional selection acting on favourable alleles in the presence of drift generates negative disequilibrium, on average, because selection rapidly eliminates positive disequilibrium whenever it arises by chance⁶. This negative disequilibrium favours the spread of a modifier allele that increases recombination, because the modifier allele is more likely to occur on chromosomes containing multiple beneficial alleles that have been brought together by past recombination^{2,6}. This advantage

to a recombination modifier can be substantial in small populations¹² or in large populations that are spatially structured¹³ and/or subject to directional selection at multiple loci¹⁴, but it requires a high rate of beneficial sweeps, for which evidence is equivocal^{15,16}. Other mechanisms that can favour recombination in large populations require certain forms of epistasis or fluctuating epistasis. For these alternative mechanisms to work, epistasis must be weak, synergistic, and similar between pairs of loci^{17,18} or it must fluctuate rapidly^{17,19}. There is currently little empirical support for such forms of epistasis^{19,20}.

Sex and recombination can also be favoured because they increase the rate of elimination of deleterious mutations. Deleterious mutations of small effect are a ubiquitous feature of living systems, and comparative molecular evolutionary analysis²¹ and experiments involving the molecular analysis of mutation accumulation lines^{22,23} suggest that there is at least one deleterious mutation per diploid per generation in taxa as diverse as *Caenorhabditis elegans*, *Drosophila*, mammals and birds. In the face of such recurrent deleterious mutations, sex and recombination can be advantageous for two distinct reasons. First, there is a deterministic advantage of recombination that eliminates the negative linkage disequilibrium generated by synergistic epistasis^{17,18,24,25}, but, as mentioned above, there is little evidence for such epistasis²⁰. Second, there is a stochastic advantage in that recombination can reduce the fixation of harmful mutations by means of Muller's ratchet⁴. Yet harmful mutations are unlikely to become fixed in sexual populations unless the effective population size is very small⁷. Even if harmful alleles do not become fixed, they can still reduce the efficacy of selection on neighbouring loci through a process called Hill–Robertson interference⁸. This effect occurs because individuals bearing deleterious mutations are less likely to survive and reproduce, reducing the number of individuals that contribute genetically to the future population. This reduces the effective population size witnessed by a focal locus, thereby increasing the importance of random genetic drift at the locus relative to selection. The extent to which the Hill–Robertson effect favours the evolution of sex and recombination in the presence of recurrent deleterious mutations is not well understood. Although previous simulations^{2,26} indicate that modifier alleles that increase the frequency of recombination can be favoured in small populations (1,000 or fewer individuals), whether this effect is appreciable in larger populations or in the presence of epistasis is unknown. To address this issue, we quantified rates of fixation of modifier alleles that increase recombination in chromosomes subject to recurrent deleterious mutations at many loci, exploring a range of population sizes and forms of epistasis.

In simulations, we allowed frequencies of deleterious mutant alleles and linkage disequilibria among them to approach mutation–selection–drift balance in a population of N haploids. We then introduced, at a random position and in a single copy, a

¹Institute of Evolutionary Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. ²Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada.

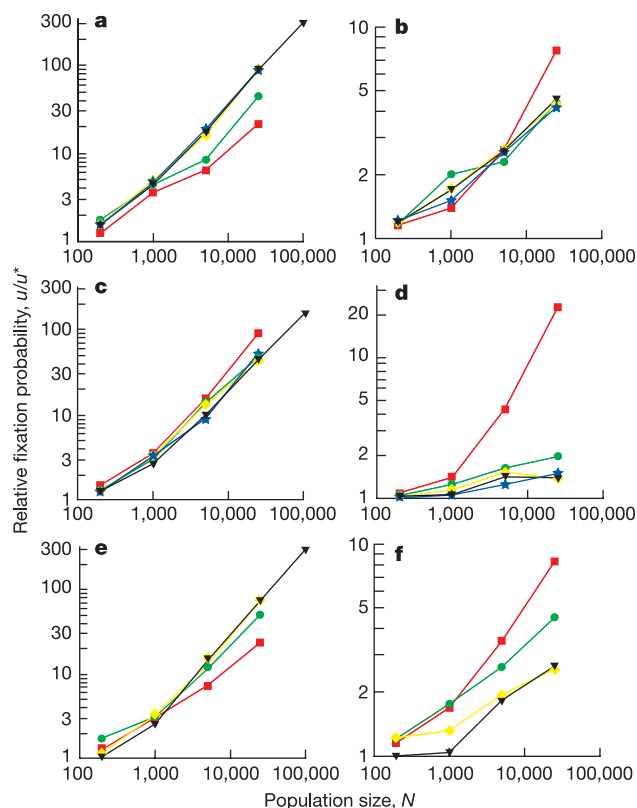


Figure 1 | Effect of model parameters on flux of modifiers of recombination. The fixation probability, u , of a modifier mutation that increases the length of a chromosome by 0.1 M, relative to the fixation probability of a neutral locus, u^* , is shown as a function of population size N , for cases of chromosomes of initial map length 0 M (a, c, e) or 0.1 M (b, d, f). Results are shown for simulations with five different values of the epistasis parameter β : 0.000001 (weak antagonistic epistasis; blue stars), 0 (black triangles), -0.000001 (yellow diamonds), -0.0001 (green circles) and -0.01 (very strong synergistic epistasis; red squares). We excluded cases in which antagonistic epistasis led to individuals with a fitness of more than 1, implying that advantageous mutations were present. The chromosomal mutation rate and fitness parameters were $U = 1$, $\alpha = 0.01$ (a, b); $U = 0.1$, $\alpha = 0.01$ (c, d); $U = 1$, $\alpha = 0.001$ (e, f). The s.e.m. of each point, measured as a percentage of the point's value, was less than or equal to 13%.

recombination modifier allele that uniformly stretched the genetic map length of a chromosome from an initial length L ($L = 0, 0.1$, or 1 morgan) to $L + 0.1$ morgan (M). The fixation probability, u , of the modifier allele was measured relative to the fixation probability, $u^* = 1/N$, of a neutral mutation. The quantity u/u^* when multiplied by the mutation rate to new modifier alleles, μ , also describes the expected rate of substitution (or 'flux') at modifier loci ($\mu Nu = \mu u/u^*$). The effect of population size on u/u^* is shown in Fig. 1 for values of the chromosomal mutation rate, U , and deleterious mutational effect of single mutations, α , that span the range of empirical estimates for several eukaryotes²⁷. We explored whether curvature in the fitness surface, as measured by the epistasis parameter β , had a major effect on the spread of the modifier (see Methods). In all simulations in which recombination started at a low level (that is, $L \leq 0.1$ M), u/u^* was greater than unity, implying that the recombination modifier is favoured. Even with chromosomes 1 M in length, the modifier of recombination tends to be favoured (Supplementary Table 1). Epistasis did not have a substantial effect on the outcome unless it was very strong and negative. For the cases considered in Fig. 1, the input of mutational variation for fitness per generation ranges from $V_M \approx U\alpha^2 = 10^{-6}$ to 10^{-4} , assuming that each mutation has an independent effect on fitness ($\beta = 0$). These values are comparable to empirical estimates of V_M per chromosome in

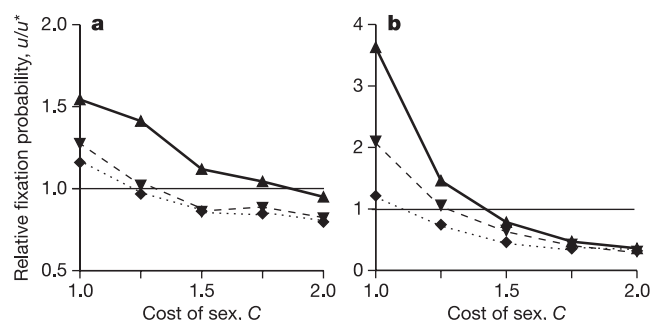


Figure 2 | Effect of model parameters on flux of modifiers of sex. The fixation probability of a modifier that causes an individual to undergo sexual reproduction with probability p_{sex} is plotted against the cost of sex, C , relative to the case of a mutation that has no effect on sex. The advantages of sex are greater than the costs for points above the horizontal line ($u/u^* = 1$). Results for two values of p_{sex} are shown: 0.01 (a) and 0.05 (b). Each simulated individual carried a single chromosome of map length 1 M with $U = 1$, $\alpha = 0.01$, no epistasis, and no sex among the remaining individuals. Solid lines, $N = 25,000$; dashed lines, $N = 5,000$; dotted lines, $N = 1,000$. The s.e.m. of each point, measured as a percentage of the point's value, was less than or equal to 9% in a and less than or equal to 14% in b.

Drosophila—typically of the order of 2×10^{-5} (ref. 28). We also investigated simulations in which a modifier mutation arose at a random location in a genome containing five chromosomes. The advantage of the recombination modifier is only slightly smaller than that observed in a single chromosome genome (Supplementary Table 2).

The flux of modifier alleles increases with the population size over the range of parameters considered (Fig. 1, and Supplementary Fig. 1), showing that Hill–Robertson interference is relevant to the evolution of sex and recombination even in large populations (for example, u/u^* can exceed 100 in populations of size 100,000). It is striking that the stochastic advantage to sex and recombination is stronger in larger populations; this surprising observation stems from the fact that such populations maintain more polymorphic loci, increasing the strength of Hill–Robertson interference.

To investigate whether Hill–Robertson interference is strong enough to favour sexual reproduction, we asked whether sex could invade an asexual population, allowing for substantial costs of sex. After the allele frequency distribution had reached steady state in the asexual population, a mutation causing individuals to undergo sexual reproduction with probability p_{sex} was introduced into the population. The resulting zygotes had one recombination event, on average, per chromosome. We introduced a cost of sex, C , reducing the number of offspring per parent when reproduction was sexual rather than asexual, namely $w_{\text{sex}} = w_{\text{asex}}/C$; we explored values of C between 1 (no cost) and 2 (a twofold cost of sex). The results (Fig. 2) indicated that selection against deleterious mutations can favour costly sex (u/u^* exceeded 1 for C values ranging from 1 to 1.75), especially when the modifier causes only a small increase in the probability of sexual reproduction. In particular, costly sex is more likely to spread in large populations.

Theory for infinite populations predicts no advantage to a modifier of recombination when genes affect fitness independently ($\beta = 0$), because such selection does not generate disequilibria¹⁰. Over the range of population sizes explored, however, our simulations show that the relative fixation probability of a modifier of recombination increases with population size under a range of models of gene action, regardless of the presence or sign of epistasis. That the Hill–Robertson effect is behind this result was confirmed by a marked decrease in the effective population size (N_e) in our simulations in comparison with the actual population size. To estimate N_e , we measured the steady-state variance at a linked locus subject to recurrent neutral mutation (see Methods). The

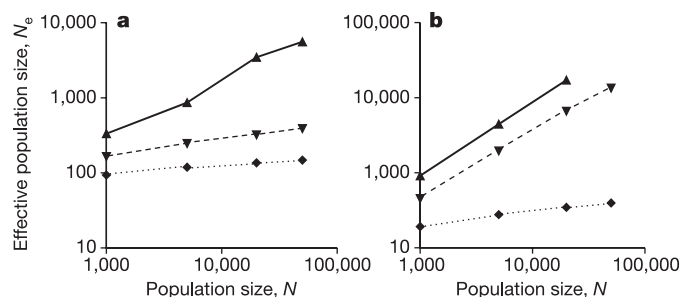


Figure 3 | Effect of background selection on effective population size. Effective population size was estimated in our simulations with $U = 1$ (a) and $U = 0.1$ (b), both with $\alpha = 0.01$ and $\beta = 0$ (no epistasis). Individuals carried only one chromosome, with the neutral locus situated at the very end. Even greater decreases in N_e are observed at centrally located neutral loci (see Supplementary Table 3). Solid lines, $L = 1$; dashed lines, $L = 0.1$; dotted lines, $L = 0$. The s.e.m. of each point, measured as a percentage of the point's value, was less than or equal to 7% in a and less than or equal to 21% in b.

results (Fig. 3, and Supplementary Table 3) show that N_e remains far below the actual population size, especially when linkage is tight. For example, in the case of complete linkage, the effective population size remains below 200 even as the census size rises to 50,000 (Fig. 3a). The decrease in effective population size observed in our simulations matched theoretical predictions²⁹ (Supplementary Table 3), and the effect of recombination on N_e was a good predictor of the fixation probability of modifier alleles (Supplementary Fig. 2). As a result of small effective population sizes, deleterious mutations became fixed in several of the simulations (even in the presence of recombination). However, such fixation events were not solely responsible for our results, because sex and recombination was also favoured in cases in which deleterious mutations rarely, if ever, became fixed (for example, $L = 0.1$ M, $U = 0.1$).

Although not directly under selection, modifiers increasing the amount of recombination can experience substantial indirect selection through associations with fitter alleles at linked loci. The strength of selection, s , can be estimated from the fixation probability by using the standard diffusion result³⁰

$$u \approx (1 - \exp[-2sN_e/N]) / (1 - \exp[-2sN_e])$$

For example, in Fig. 1a ($L = 0$ M, $U = 1$, $\alpha = 0.01$, no epistasis), selection acting on the modifier equals $s \approx 0.02$ when $N = 1,000$ and rises to $s \approx 0.5$ when $N = 50,000$, whereas in Fig. 1b ($L = 0.1$ M), selection rises from $s \approx 0.004$ when $N = 1,000$ to $s \approx 0.008$ when $N = 50,000$ (using values of N_e estimated in our simulations). Drift hinders the fixation of even very beneficial modifier alleles because of the drastic decrease in the effective size relative to the census size of the population.

There are two principal conclusions from our results. First, theory developed for infinite populations to predict the fate of a modifier does not apply, because random genetic drift in the presence of selection generates disequilibria that can favour a modifier, even in large populations. Second, over the range of parameters explored, Hill–Robertson effects overwhelm the effects of epistasis as a force generating disequilibria among alleles at different loci. Consequently, we find that the form of epistasis is not critical to the advantage of sex and recombination in finite populations, in contrast to theoretical predictions from infinite populations at mutation–selection balance^{17,18,24,25}. These conclusions apply over a wide range of plausible values for the genomic deleterious mutation rate and mean effect of deleterious mutations.

Although we have focused on the fixation probability of a single modifier mutation, simulations allowing recurrent modifier mutations show that chromosomal recombination rates can rise to the order of 1 M (Fig. 4). Multi-locus Hill–Robertson interference

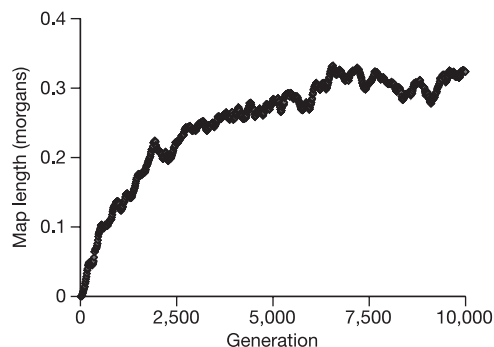


Figure 4 | The long-term evolution of recombination allowing recurrent mutation at modifier loci. The mean map length from 10 replicate simulations is plotted against generation number. Background selection was simulated on a single chromosome without epistasis ($U = 1$, $\alpha = 0.01$, $\beta = 0$). Recurrent recombination modifier mutations arose at 100 random loci on the chromosome at a rate of 0.001 per chromosome. Their effects were $+0.1$ M or -0.1 M, with equal probability. The population size was 1,000.

therefore provides a general and robust explanation for the evolution of recombination in any genome subject to recurrent deleterious mutations and can even contribute to the evolution of costly sex.

METHODS

We simulated haploid populations of N individuals with genomes of c independently segregating chromosomes, each containing 100 equally spaced loci affecting fitness. In each generation, deleterious mutations affecting fitness occurred before mating and zygote formation. The number of mutations per generation per individual was sampled from a Poisson distribution with a mean of U . Each mutation was assigned to a random locus in the genome. We kept track of the number of mutations carried at each locus, so that we could allow for multiple mutations. The fitness of an individual i was $w_i = \exp[-\alpha n_i + \beta n_i^2]$, where α is the independent fitness effect of a mutation, β its epistatic effect, and n_i the number of deleterious mutations carried by the individual. To ensure that mutations were always deleterious, we could explore only a narrow range of positive values of β . To form a mating pair, individuals were sampled with replacement with probability proportional to w_i . Having selected a mating pair, a zygote was formed by allowing n_c recombination events to occur between the pair's chromosomes, where n_c was sampled from a Poisson distribution with mean L to simulate a chromosome of initial map length L Morgans. These n_c recombination events were then randomly and uniformly distributed between the 100 loci. This process was repeated until N offspring were produced, leading to a nearly Poisson distribution of offspring per parent.

To simulate the fate of a recombination modifier, the allele frequencies at the loci affecting fitness were allowed to approach their steady-state frequencies by allowing a long 'burn-in' period of at least N generations of mutation, selection, drift and recombination. This burn-in period was set to many times the effective population size measured in the simulations (see below). After the burn-in, the state of the population was saved. A recombination modifier mutation was introduced to a random individual of the burn-in population at a random position on a chromosome, coinciding with one of the fitness-altering mutations. The recombination modifier increased the expected number of recombination events in zygotes heterozygous (homozygous) for the modifier to $L + 0.05$ M ($L + 0.1$ M). For each burn-in population, the fates of between N and $5N$ modifier mutations were tracked. At least five replicate burn-in populations were simulated for each parameter combination. The fraction of recombination modifiers that became fixed per burn-in population was recorded, and the mean and s.e.m. of this fraction were calculated over independent burn-ins.

The simulation of the fate of a modifier of sex was similar to that for a modifier of recombination. Asexual burn-in populations were simulated, and then a mutation was introduced in a single copy that caused its haploid carrier to produce all of its offspring sexually with probability p_{sex} but at a cost, C .

To estimate effective population size (N_e) in the presence of deleterious mutations at linked loci, a neutral, linked locus was incorporated in the simulation. This locus was either telomeric or centrally located on the chromosome. In each generation, the value of each individual's neutral locus was altered by adding a normally distributed mutational effect of mean 0 and variance

$V_M = 1$. In the absence of selection, the equilibrium variance at the locus is expected to be NV_M , and this was confirmed in simulations. With selection at linked loci, the mean equilibrium variance at the neutral locus was then defined as the effective population size, N_e . A burn-in period of at least seven times the equilibrium N_e computed from theory was allowed; N_e was then computed in each generation for at least a further 10N generations. Average N_e estimates from independent burn-ins were used to calculate an overall estimate of the mean and standard error of N_e .

Received 19 May; accepted 5 July 2006.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, 1930).
2. Felsenstein, J. & Yokoyama, S. The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845–859 (1976).
3. Morgan, T. H. *Heredity and Sex* (Columbia Univ. Press, New York, 1913).
4. Muller, H. J. Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
5. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
6. Barton, N. H. & Otto, S. P. Evolution of recombination due to random drift. *Genetics* **169**, 2353–2370 (2005).
7. Lynch, M., Conery, J. & Bürger, R. Mutational meltdowns in sexual populations. *Evolution Int. J. Org. Evolution* **49**, 1067–1080 (1995).
8. Hill, W. G. & Robertson, A. The effects of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
9. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ. of California Press, Berkeley, 1982).
10. Felsenstein, J. The effect of linkage on directional selection. *Genetics* **52**, 349–363 (1965).
11. Bürger, R. Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* **153**, 1055–1069 (1999).
12. Otto, S. P. & Barton, N. H. Selection for recombination in small populations. *Evolution Int. J. Org. Evolution* **55**, 1921–1931 (2001).
13. Martin, G., Otto, S. P. & Lenormand, T. Selection for recombination in structured populations. *Genetics* **172**, 593–609 (2006).
14. Iles, M. M., Walters, K. & Cannings, C. Recombination can evolve in large finite populations given selection on sufficient loci. *Genetics* **165**, 333–337 (2003).
15. Bierne, N. & Eyre-Walker, A. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol. Biol. Evol.* **21**, 1350–1360 (2004).
16. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
17. Barton, N. H. A general model for the evolution of recombination. *Genet. Res.* **65**, 123–144 (1995).
18. Otto, S. P. & Feldman, M. W. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* **51**, 134–147 (1997).
19. Otto, S. P. & Nuismer, S. Species interactions and the evolution of sex. *Science* **304**, 1018–1020 (2004).
20. Rice, W. R. Experimental tests of the adaptive significance of sexual recombination. *Nature Rev. Genet.* **3**, 241–251 (2002).
21. Keightley, P. D. & Eyre-Walker, A. Deleterious mutations and the evolution of sex. *Science* **290**, 331–333 (2000).
22. Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682 (2004).
23. Halligan, D. L. & Keightley, P. D. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res.* **16**, 875–884 (2006).
24. Kondrashov, A. S. Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet. Res.* **44**, 199–217 (1984).
25. Charlesworth, B. Mutation selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221 (1990).
26. Pálsson, S. Selection on a modifier of recombination rate due to linked deleterious mutations. *J. Hered.* **93**, 22–26 (2002).
27. Charlesworth, B. & Charlesworth, D. Some evolutionary consequences of deleterious mutations. *Genetica* **103**, 3–19 (1998).
28. Caballero, A. & Keightley, P. D. A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900 (1994).
29. Santiago, E. & Caballero, A. Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* **149**, 2105–2117 (1998).
30. Kimura, M. Diffusion models in population genetics. *J. Appl. Probab.* **1**, 177–232 (1964).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Santiago and A. Caballero for advice on predicting effective population size; N. Barton, A. Blachford, C. Haag, W. Hill, D. Roze and M. Whitlock for comments on the manuscript; and M. Blaxter for Linux cluster computing time. S.P.O. was supported by a NSERC grant (Canada).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.K. (keightley.nature2006@spambob.net).

Evidence for complete denitrification in a benthic foraminifer

Nils Risgaard-Petersen¹, Alexandra M. Langezaal², Signe Ingvarsdén³, Markus C. Schmid⁴, Mike S. M. Jetten⁴, Huub J. M. Op den Camp⁴, Jan W. M. Derksen⁵, Elisa Piña-Ochoa⁶, Susanne P. Eriksson⁷, Lars Peter Nielsen³, Niels Peter Revsbech³, Tomas Cedhagen³ & Gijsbert J. van der Zwaan²

Benthic foraminifera are unicellular eukaryotes found abundantly in many types of marine sediments. Many species survive and possibly reproduce in anoxic habitats¹, but sustainable anaerobic metabolism has not been previously described. Here we demonstrate that the foraminifer *Globobulimina pseudospinescens* accumulates intracellular nitrate stores and that these can be respired to dinitrogen gas. The amounts of nitrate detected are estimated to be sufficient to support respiration for over a month. In a Swedish fjord sediment where *G. pseudospinescens* is the dominant foraminifer, the intracellular nitrate pool in this species accounted for 20% of the large, cell-bound, nitrate pool present in an oxygen-free zone. Similarly high nitrate concentrations were also detected in foraminifera *Nonionella cf. stella* and a *Stainforthia* species, the two dominant benthic taxa occurring within the oxygen minimum zone of the continental shelf off Chile. Given the high abundance of foraminifera in anoxic marine environments^{1–3}, these new findings suggest that foraminifera may play an important role in global nitrogen cycling and indicate that our understanding of the complexity of the marine nitrogen cycle is far from complete.

In 2005 we investigated foraminifera in the oxygen-free zone of sediment from the Gullmar Fjord, Sweden. By promoting cell lysis we identified a large nitrate pool beneath the oxygenated zone at a depth where the sediment pore water is devoid of nitrate (Fig. 1). This cell-bound nitrate pool positively correlated with foraminiferal abundance (Pearson correlation, $r = 0.98$; $P = 0.0001$, $n = 7$). One of the dominant foraminifera in this sediment was *G. pseudospinescens* (Fig. 2a), which accounted for 37% of the standing stock in the $>150\ \mu\text{m}$ fraction. As is typical for this organism⁴, it mainly occurred in sediment layers where the pore water lacked oxygen and nitrate (Fig. 1). The average nitrate content of individual *G. pseudospinescens* was 18 nmol (range 0–72 nmol; Table 1), corresponding to an average intracellular concentration well above 10 mM, or more than 500-fold higher than the maximum concentration in the sediment pore water. The nitrate in *G. pseudospinescens* accounted for about 20% of the total cell-bound nitrate pool in the sediment, thus leaving open the possibility that other foraminifera also accumulate nitrate in this sediment. Possible candidates include *Nonionella turgida* and *Stainforthia fusiformis*, which also occurred in anoxic layers of the sediment but were not analysed for nitrate.

That accumulation of nitrate in foraminifera might be a widespread phenomenon was confirmed through a study of the epibenthic community within the oxygen-minimum zone of the continental shelf off Chile in January 2006. The intracellular nitrate

concentration in the two dominant taxa—*Nonionella cf. stella* (Fig. 2b) and an undescribed *Stainforthia* species (Fig. 2c)—was even higher than in *G. pseudospinescens* and up to 15,000-fold higher than the maximum porewater concentration (Table 1).

To accumulate nitrate against these large electrochemical gradients requires energy⁵ and hence is only sustainable if the foraminifera benefit from their intracellular nitrate pool. It has been previously proposed that species preferring anoxic environments can respire nitrate⁶. We therefore tested whether the intracellular nitrate pool in *G. pseudospinescens* is denitrified (that is, respired).

When individual *G. pseudospinescens* were incubated in oxygen-free water containing $^{15}\text{NO}_3^-$, ^{15}N -labelled dinitrogen gas was formed, thus indicating complete denitrification (that is, full reduction of nitrate to dinitrogen gas (one-tailed t -test, $P = 0.0005$, $n = 8$, $\alpha = 0.05$; Fig. 3a). The following observations

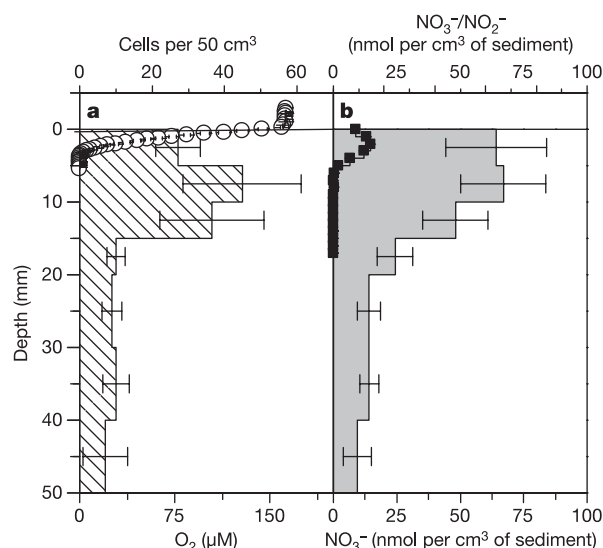


Figure 1 | Foraminifera, oxygen and nitrate in sediment. **a**, Depth distribution of Rose-Bengal stained *G. pseudospinescens* (hatched area) and oxygen (open circles). **b**, Porewater nitrate/nitrite (squares) and cell-bound nitrate (grey area) in sediment cores from the Gullmar Fjord, Sweden, August 2005. The error bars indicate the s.e.m. ($n = 3$).

¹Department of Marine Ecology, National Environmental Research Institute, Vejløvej 25, DK-8600 Silkeborg, Denmark. ²Department of Earth Sciences, Utrecht University, Faculty of Geosciences, Budapestlaan 4, NL-3584 CD Utrecht, The Netherlands. ³Department of Biological Sciences, University of Aarhus, Ny Munkegade, Building 1540, DK-8000 Aarhus C, Denmark. ⁴Department of Microbiology and ⁵Department of Plant Cell Biology, IWWIR, Radboud University Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands. ⁶Department of Aquatic Ecology (CSIC), Centre of Environmental Sciences, Serrano 115 dpdo., E-28006 Madrid, Spain. ⁷Department of Marine Ecology, Kristineberg Marine Research Station, Göteborg University, S-450 34 Fiskebäckskil, Sweden.

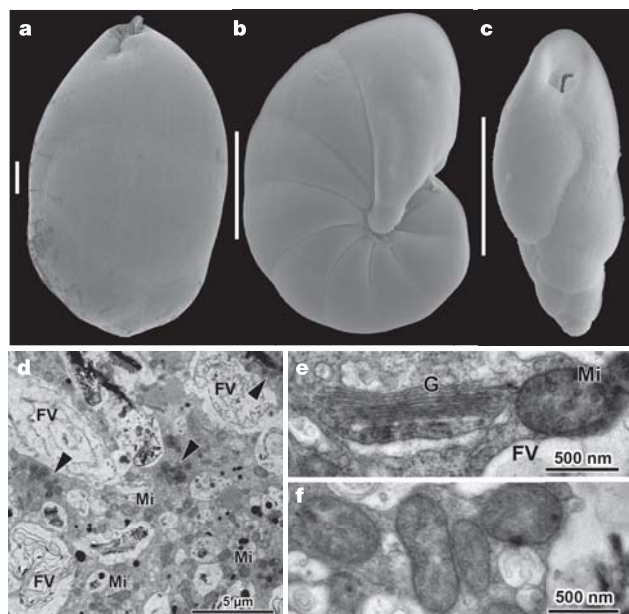


Figure 2 | SEM and TEM micrographs of the investigated foraminifera species. **a–c**, SEM micrographs of single cells of *G. pseudospinescens* (**a**), *Nonionella cf. stella* (**b**) and *Stainforthia* sp. (**c**). Scale bar, 100 μm . **d**, TEM micrograph showing cellular organization of *G. pseudospinescens* containing the intact cytoplasm with multiple food vacuoles (FV) and clusters of mitochondria (Mi) and peroxisomes (arrows). **e**, Detailed TEM micrograph showing a mitochondrion (Mi) and Golgi apparatus (G). **f**, Detailed TEM micrograph of a cluster of mitochondria.

show that this process took place intracellularly using the intracellular nitrate ($^{14}\text{NO}_3^-$) as the substrate: if denitrification had only taken place outside the foraminifera, less than 2% of the recovered ^{15}N would have appeared as mixed-labelled $^{29}\text{N}_2$ because the ^{15}N -labelling of nitrate in the medium exceeded 98 atom% throughout the incubation. The appearance of recovered ^{15}N in the $^{29}\text{N}_2$ pool was significantly different from (two-tailed t -test, $P = 0.0037$, $n = 8$, $\alpha = 0.05$) and generally much higher than this hypothetical fraction. In 6 out of 8 samples, more than 40% of the recovered ^{15}N appeared as the mixed-labelled $^{29}\text{N}_2$. As parallel experiments with $^{15}\text{NH}_4^+$ and $^{14}\text{NO}_3^-$ excluded the possibility of anaerobic ammonium oxidation (that is, the anammox process⁷), $^{29}\text{N}_2$ production must have been due solely to denitrification. The foraminifera must therefore have taken up $^{15}\text{NO}_3^-$ from the medium, mixed this with their intracellular pool of $^{14}\text{NO}_3^-$ and denitrified the mixed $^{14}\text{NO}_3^-/^{15}\text{NO}_3^-$ pool intracellularly.

In a second series of experiments, the denitrification rate was measured in individual *G. pseudospinescens* containing $^{14}\text{NO}_3^-$ and $^{15}\text{NO}_3^-$. The specimens were prepared by incubating intact sediment cores with 50 μM $^{15}\text{NO}_3^-$ in the overlying water. Foraminifera harvested from the cores after 8 weeks exhibited an average

($\pm$ s.e.m.) of 40 ± 4 ^{15}N atom% enrichment of their nitrate pool. Nineteen of the twenty-two cells analysed contained measurable $^{15}\text{NO}_3^-$. Denitrification was measured on foraminifera harvested after 9 weeks and incubated in nitrate-free and oxygen-free media. In 10 out of 12 cases, $^{29}\text{N}_2$ and $^{30}\text{N}_2$ production was detected during a 24–72-h incubation period (Fig. 3b). Linear regression analysis revealed an average dinitrogen production rate for a foraminifer cell of 565 pmol N d^{-1} with a 95% confidence interval between 226–902 pmol N d^{-1} .

In subsequent experiments, evidence for complete denitrification of intracellular nitrate was also obtained for *Nonionella cf. stella*. Specimens were incubated in microtubes containing nitrate-free and oxygen-free water in the presence or absence of acetylene, which inhibits the biological reduction of nitrous oxide to dinitrogen. Denitrification was quantified from the accumulation of nitrous oxide (see Supplementary Information for details). Nitrous oxide only accumulated in the presence of acetylene, corresponding to a denitrification rate for a single cell of 84 ± 33 pmol N d^{-1} ($\pm$ s.e.m.).

Using two different methodologies we have thus shown that the intracellular nitrate in the foraminifera is denitrified. Little is known about the physiology of foraminifera, however, and the denitrification rates *in situ* may well be significantly higher or lower than the rates reported here based on laboratory incubations.

There are two possible mediators of the observed denitrification inside foraminifera: endosymbiotic nitrate-reducing bacteria or denitrifying enzymes in the foraminifera's own organelles.

As an efficient denitrifying prokaryote has a specific denitrification capacity of about 40 $\text{fmol N cell}^{-1} \text{d}^{-1}$ (see Supplementary Information⁸), the denitrification rate measured in *G. pseudospinescens* would necessitate the presence of between 5,650 and 22,550 endosymbiotic denitrifying bacteria in each cell. To quantify bacteria that might be involved in nitrate respiration we analysed *G. pseudospinescens* by fluorescence *in situ* hybridization (FISH) and scanning electron microscopy (SEM). In addition, we examined its cellular ultrastructure by transmission electron microscopy (TEM) of thin sections. Analysis of individual *G. pseudospinescens* with 4',6-diamidino-2-phenylindole (DAPI) and FISH probes for members of Archaea or Bacteria revealed less than 100 prokaryotic cells on the surface of each specimen. Crushing the foraminifera did not release more prokaryotic cells. Sediment samples from the same location used as controls for the FISH analyses contained $2.7 \pm 0.4 \times 10^9$ bacterial cells per ml of sediment. Examination of the foraminifera by SEM and TEM supported the FISH results: there were very few prokaryotic cells attached to their surface and no endosymbionts present. The TEM micrographs of fully sectioned cells revealed an intact cytoplasm with mitochondria and other organelles. No bacterial infestation was visible other than occasional occurrences within food vacuoles. These findings therefore strongly suggest that denitrification should be ascribed to the foraminifera's own organelles, for instance, their mitochondria.

Non-symbiotic nitrate respiration is known to occur in three eukaryotes—the anaerobic freshwater protozoa *Loxodes* spp., which only reduces nitrate as far as nitrite⁹, and the fungi *Fusarium*

Table 1 | Intracellular nitrate content and concentration in foraminifera

Location	Maximum NO_3^- concentration in pore water (μM)	Species	Intracellular NO_3^- content (pmol per cell)		Intracellular NO_3^- concentration (mM)
			Mean	Range	
Gullmar Fjord, Sweden	17	<i>G. pseudospinescens</i>	18,000 (5,000, $n = 20$)	0–72,000	10 (2.3)
Station 18, Chile	12	<i>Nonionella cf. stella</i>	186 (25, $n = 43$)	8–794	35
Station 18, Chile	12	<i>Stainforthia</i> sp.	60 (9, $n = 26$)	0–172	180

The maximum ambient porewater nitrate concentration is shown for comparison. The s.e.m. and number of replicates are given in parentheses. The nitrate content of approximately 2 mm^3 control samples of sediment particles from the sites was less than 20 pmol in the Swedish sediment (that is, close to the detection limit for the method used at this site) and 4.6 ± 1.5 pmol in the Chilean sediment. Note that the intracellular nitrate concentrations should be considered minimum values as the estimates are based on shell volume, rather than cell volume or the volume of possible storage vacuoles. Shell volumes were as follows: *G. pseudospinescens*, $0.5\text{--}3.6 \text{ mm}^3$; *Nonionella cf. stella*, $5.2 \times 10^{-3} \pm 7.1 \times 10^{-4} \text{ mm}^3$; *Stainforthia* sp., $3.3 \times 10^{-4} \pm 2.06 \times 10^{-5} \text{ mm}^3$.

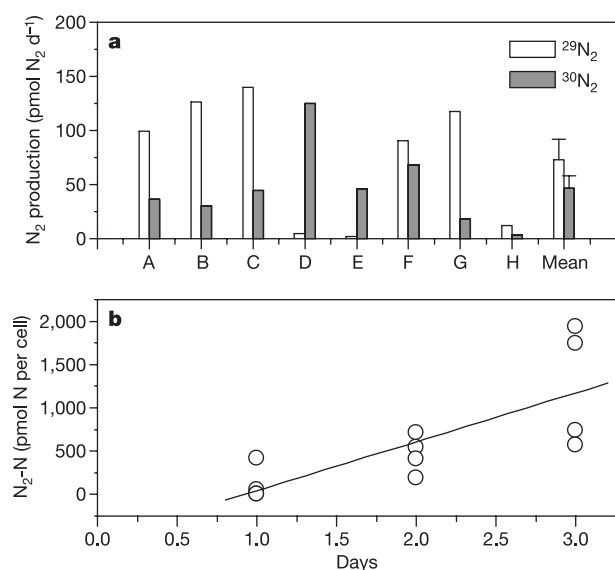


Figure 3 | N_2 production in *G. pseudospinescens*. **a**, $^{29}N_2$ and $^{30}N_2$ production in samples each containing two *G. pseudospinescens* cells and $^{15}NO_3^-$ (98.0 ± 0.05 ^{15}N atom%). Letters A–H, individual samples. Mean, values averaged across all samples. The error bars indicate the s.e.m. ($n = 8$). **b**, N_2-N (where N_2-N is dinitrogen expressed in units of N atoms) produced by individual *G. pseudospinescens* containing a ^{15}N -labelled intracellular nitrate pool. Circles represent discrete samples each containing two cells of *G. pseudospinescens*. The amount of N_2-N produced was calculated from the measured production of $^{29}N_2$ and $^{30}N_2$ during the incubations using the isotope pairing technique²⁶. The regression line is shown ($R^2 = 0.58$, $P < 0.01$, $n = 12$).

oxyphorum and *Cylindrocarpon tonkinense*¹⁰, which reduce nitrate to nitrous oxide. Complete denitrification such as observed here for foraminifera has not been previously reported for eukaryotic cells.

A specific advantage of using nitrate for respiration is that it is possible to accumulate this non-toxic ion to high concentrations in the cell, thereby enabling respiration to be sustained during periods when nitrate is absent from the ambient environment. Assuming that the denitrification rate estimated for single *G. pseudospinescens* cells (that is, $565 \text{ pmol N d}^{-1}$) is representative for the organisms in their natural habitat, they would thus be able to survive in the absence of external nitrate sources for over a month using their intracellular nitrate pool of about 18 nmol NO_3^- (Table 1). This would explain their ability to remain in sediment layers lacking ambient oxygen and nitrate^{4,11}. As *Globobulimina* is known to be motile and can also be found at the surface of sediments^{4,11}, the organism could replenish its nitrate pool there before migrating to deeper anoxic strata where competition and predation pressure are minimal. Temporary nitrate accumulation by foraminifera may also be beneficial in the oxygen-minimum zone of the continental shelf off Chile, where nitrate uptake by filamentous sulphur bacteria restricts the occurrence of nitrate to the sediment surface¹². Given their nitrate storage and denitrification capabilities, *Nonionella* cf. *stella* could make short foraging trips into organic-matter-rich, nitrate-free sediment strata.

The present finding of combined nitrate accumulation and complete denitrification within eukaryotic cells indicates that the present conceptual view of marine denitrification as a bacteria-mediated process is incomplete. Denitrification of intracellular nitrate in motile foraminifera represents a hitherto unknown pathway for removal of nitrogen from the marine environment. Future studies addressing the generality among members of the Foraminifera of the traits discussed here will provide valuable information about the evolutionary origin of the trait and the quantitative importance of this pathway for global nitrogen cycling.

METHODS

Sampling was performed in Gullmar Fjord, Sweden ($58^\circ 19.4' N$; $11^\circ 32.7' E$; Station S3 in ref. 13) in August 2005 and at Station 18 off Concepcion Bay, Chile ($36^\circ 32.1' S$; $73^\circ 7.1' W$)¹⁴ in January 2006. The bottom water concentrations of oxygen and nitrate were $160 \mu M$ and $15 \mu M$, respectively, in Gullmar Fjord (temperature $6^\circ C$) and $<1 \mu M$ and $12 \mu M$, respectively, off Concepcion Bay (temperature $11^\circ C$).

Nitrate/nitrite, and oxygen depth profiles ($n = 3-5$) were measured in sediment cores from Gullmar Fjord using a biosensor¹⁵ and a Clark-type oxygen microsensor¹⁶ as previously described¹⁷. The measurements were performed at $10^\circ C$, which is the lower limit of the biosensor's operational range. The cell-bound nitrate pool was quantified in three cores using a described approach¹⁸, except that cell lysis was promoted by placing sediment slices in centrifuge tubes and immersing them in boiling water for 10 min.

Foraminifera for nitrate measurements were collected from the 0.5–2-cm core depth interval (Swedish sediment) or from the surface of the cores (Chilean sediment). They were gently cleaned with a brush and kept overnight in Petri dishes containing sea water. Only specimens showing renewed collection of material at the aperture were used for the measurements.

The nitrate content of individual *G. pseudospinescens* was measured using the VCl_3 reduction method¹⁹ on a chemiluminescence detector (Model 42; Thermo Environmental Instruments). Before nitrate determination each foraminifer was measured for estimation of shell volume, rinsed in nitrate-free artificial sea water, dried, decalcified for 1 h at $4^\circ C$ in $100 \mu l$ of IUPAC buffer solution (pH 4; Radiometer analytical) and then crushed. Nitrate was measured in 12–25- μl subsamples. The specimens of *Nonionella* cf. *stella* and *Stainforthia* sp. were inserted directly into 0.6 ml of $2 M$ HCl solution with VCl_3 held at $80^\circ C$ in a reduction chamber connected to the detector (Model CLD 86; Eco Physics AG). Their shell volume was estimated from the linear dimensions of a separate batch of specimens. The nitrate content of sediment samples (approximately 2 mm^3 ; $n = 11$ from each site) was determined following the same protocol as for the foraminifera. Intracellular nitrate concentration was estimated from the measured nitrate content and the estimated shell volume.

Quantitative FISH with the probe SD-Arch-0915-aA-20 specific for Archaea, the EUBmix (SD-Bact-0338-aA-18, S-*BactP-0338-aA-18, S-*BactV-0338-aA-18) specific for Bacteria (Thermo Electron Corporation), and the general DNA stain (DAPI) was performed on 12 *G. pseudospinescens*; all appropriate controls being included²⁰. Control hybridization studies of sieved sediment from the same sampling confirmed that bacteria could be identified and quantified by FISH if present. SEM and TEM were performed on six *G. pseudospinescens* as previously described²¹. One complete cell was sectioned at $0.5-1 \mu m$ intervals and examined by light microscopy. From these sections, around 100 randomly selected ultra-thin sections (100 nm) were used for TEM of the cytoplasm.

Denitrification in *G. pseudospinescens* was measured using ^{15}N methods in two experimental rounds. Round one: two series of 6-ml glass vials (Exetainer; Labco) each containing two cells in filter-sterilized ($0.2 \mu m$; Sartorius) oxygen-free, He-flushed medium (based on artificial sea water) with either $500 \mu M$ $^{15}NO_3^-$ (98 atom% Cambridge Isotope Laboratories), $500 \mu M$ acetate and $500 \mu M$ $^{14}NH_4Cl$ (denitrification medium; $n = 8$) or $500 \mu M$ $^{15}NH_4^+$ and $500 \mu M$ $^{14}NO_3^-$ (anammox medium; $n = 7$) was incubated for 24–48 h at $10^\circ C$. Thereafter 4 ml of medium was removed from the vials²² for determination of $^{15}NO_3^-$ atom%²³ and $100 \mu l$ of He-flushed $ZnCl_2$ was injected through the rubber septum of the vials to halt N_2 production^{7,22}. $^{29}N_2$ and $^{30}N_2$ were determined as described previously²².

Round two: *G. pseudospinescens* containing ^{15}N -labelled intracellular nitrate pools were prepared through incubation of intact cores in 30-l tanks for 8 or 9 weeks at $10^\circ C$ with $50 \mu M$ $^{15}NO_3^-$ in the overlying water. The 8-week incubations were used for determining the ^{15}N atom% of intracellular nitrate using a standard method²³ modified for small volumes. The 9-week incubations were used for measuring denitrification. Twelve Exetainers were prepared with 2 ml of sterile nitrate-free and oxygen-free medium based on sediment pore water²⁴ with auto-claved algal cells (*Dunaliella* sp.). Two *G. pseudospinescens* were added to each vial, the vials capped and the headspace flushed with He for 4 min. Incubation was halted at 24-h intervals within a 72-h incubation period as described above. Control experiments based on fluorescein diacetate staining²⁵ revealed 100% survival of *G. pseudospinescens* after incubation for more than 72 h under these conditions.

Denitrification in *Nonionella* cf. *stella* was measured using N_2O microelectrodes and acetylene inhibition (see Supplementary Information for details).

Received 2 June; accepted 13 July 2006.

- Bernhard, J. M. Microaerophilic and facultative anaerobic benthic foraminifera: a review of experimental and ultrastructural evidence. *Rev. de Paléobiol* 15, 261–275 (1996).

2. Gooday, A. J., Bernhard, J. M., Levin, L. A. & Suhr, S. B. Foraminifera in the Arabian Sea oxygen minimum zone and other oxygen-deficient settings: taxonomic composition, diversity, and relation to metazoan faunas. *Deep-sea Res. II* **47**, 25–54 (2000).
3. Levin, L. *et al.* Benthic processes on the Peru margin: a transect across the oxygen minimum zone during the 1997–98 El Niño. *Prog. Oceanogr.* **53**, 1–27 (2002).
4. Fontanier, C. *et al.* Seasonal and interannual variability of benthic foraminiferal faunas at 550 m depth in the Bay of Biscay. *Deep-sea Res. I* **50**, 457–494 (2003).
5. Alberts, B., *et al.* *Molecular Biology of the Cell* (Garland Science, New York, 2002).
6. Bernhard, J. M. & Reimers, C. E. Benthic foraminiferal population fluctuations related to anoxia: Santa Barbara Basin. *Biogeochemistry* **15**, 127–149 (1991).
7. Dalsgaard, T., Canfield, D. E., Petersen, J., Thamdrup, B. & Acuña-González, J. N₂ production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica. *Nature* **422**, 606–608 (2003).
8. Chen, F., Xia, Q. & Ju, L. K. Aerobic denitrification of *Pseudomonas aeruginosa* monitored by online NAD(P)H fluorescence. *Appl. Environ. Microbiol.* **69**, 6715–6722 (2003).
9. Finlay, B. J., Span, A. S. W. & Harman, J. M. P. Nitrate respiration in primitive eukaryotes. *Nature* **303**, 333–336 (1983).
10. Tielens, A. G. M., Rotte, C., van Hellemond, J. J. & Martin, W. Mitochondria as we don't know them. *Trends Biochem. Sci.* **27**, 564–572 (2002).
11. Geslin, E., Heinz, P., Jorissen, F. & Hemleben, C. Migratory responses of deep-sea benthic foraminifera to variable oxygen conditions: laboratory investigations. *Mar. Micropaleontol.* **53**, 227–243 (2004).
12. Zopfi, J., Kjaer, T., Nielsen, L. P. & Jørgensen, B. B. Ecology of *Thioploca* spp.: nitrate and sulfur storage in relation to chemical microgradients and influence of *Thioploca* spp. on the sedimentary nitrogen cycle. *Appl. Environ. Microbiol.* **67**, 5530–5537 (2001).
13. Engström, P., Dalsgaard, T., Hulth, S. & Aller, R. C. Anaerobic ammonium oxidation by nitrite (anammox): Implications for N₂ production in coastal marine sediments. *Geochim. Cosmochim. Acta* **69**, 2057–2065 (2005).
14. Huettel, M., Forster, S., Klöser, S. & Fossing, H. Vertical migration in the sediment-dwelling sulfur bacteria *Thioploca* spp. in overcoming diffusion limitations. *Appl. Environ. Microbiol.* **62**, 1863–1872 (1996).
15. Larsen, L. H., Kjaer, T. & Revsbech, N. P. A microscale NO₃⁻ biosensor for environmental applications. *Anal. Chem.* **69**, 3527–3531 (1997).
16. Revsbech, N. P. An oxygen microelectrode with a guard cathode. *Limnol. Oceanogr.* **34**, 474–478 (1989).
17. Meyer, R. L., Kjaer, T. & Revsbech, N. P. Use of NO₃⁻ microsensors to estimate the activity of sediment nitrification and NO₃⁻ consumption along an estuarine salinity, nitrate, and light gradient. *Aquat. Microb. Ecol.* **26**, 181–193 (2001).
18. Sayama, M. Presence of nitrate-accumulating sulfur bacteria and their influence on nitrogen cycling in a shallow coastal marine sediment. *Appl. Environ. Microbiol.* **67**, 3481–3487 (2001).
19. Braman, R. S. & Hendrix, S. A. Nanogram nitrite and nitrate determination in environmental and biological materials by vanadium(III) reduction with chemiluminescence detection. *Anal. Chem.* **61**, 2715–2718 (1989).
20. Kuypers, M. M. M. *et al.* Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**, 608–611 (2003).
21. Raghoebarsing, A. A. *et al.* Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature* **436**, 1153–1156 (2005).
22. Risgaard-Petersen, N. & Rysgaard, S. in *Methods in Applied Soil Microbiology and Biochemistry* (eds Alef, K. & Nannipieri, P.) 287–296 (Academic, London, 1995).
23. Risgaard-Petersen, N., Rysgaard, S. & Revsbech, N. P. A sensitive assay for determination of ¹⁴N/¹⁵N isotope distribution in NO₃⁻. *J. Microbiol. Meth.* **17**, 155–164 (1993).
24. Vester, F. & Ingvorsen, K. Improved most-probable-number method to detect sulfate-reducing bacteria with natural media and a radiotracer. *Appl. Environ. Microbiol.* **64**, 1700–1707 (1998).
25. Bernhard, J. M. Distinguishing live from dead foraminifera: methods review and proper applications. *Micropaleontology* **46**, 38–46 (2000).
26. Nielsen, L. P. Denitrification in sediments determined from nitrogen isotope pairing. *FEMS Microbiol. Ecol.* **86**, 357–362 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Engström, J. Brandsma, V. A. Gallardo, B. B. Jørgensen, E. Frandsen, G. Ittman, S. Petersen, K. L. Knudsen, G. J. Janssen, P. Sørensen and L. Pierson for their support. The work was supported by the Carlsberg Foundation, Denmark (N.R.-P.); the National Science Research Foundation, Denmark (N.R.-P., T.C. and N.P.R.); the Netherlands Organization for Scientific Research (NWO-ALW biogeosphere; G.J.v.d.Z. and M.S.M.J.) and the Swedish Research Council (S.P.E.).

Author Contributions A.M.L., M.C.S., N.R.-P., S.I., L.P.N. and N.P.R. performed the sampling. A.M.L., G.J.v.d.Z. and T.C. carried out the foraminifer identification. N.R.-P., N.P.R., E.P.-O., S.I., L.P.N. and A.M.L. performed the nitrate and denitrification rate measurements; J.W.M.D., H.J.M.O.d.C., T.C. and M.C.S. performed the microscopy. The research was conceived by N.R.-P., S.P.E., G.J.v.d.Z., M.S.M.J., H.J.M.O.d.C., A.M.L., N.P.R., L.P.N. and S.I. All authors contributed to interpreting the data and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.R.-P. (nri@dmu.dk).

Soma-germline interactions coordinate homeostasis and growth in the *Drosophila* gonad

Lilach Gilboa¹ & Ruth Lehmann¹

The ability of organs such as the liver or the lymphoid system to maintain their original size or regain it after injury is well documented^{1,2}. However, little is known about how these organs sense that equilibrium is breached, and how they cease changing when homeostasis is reached. Similarly, it remains unclear how, during normal development, different cell types within an organ coordinate their growth. Here we show that during gonad development in the fruitfly *Drosophila melanogaster* the proliferation of primordial germ cells (PGCs) and survival of the somatic intermingled cells (ICs) that contact them are coordinated by means of a feedback mechanism composed of a positive signal and a negative signal. PGCs express the EGF receptor (EGFR) ligand Spitz, which is required for IC survival. In turn, ICs inhibit PGC proliferation. Thus, homeostasis and coordination of growth between soma and germ line in the larval ovary is achieved by using a sensor of PGC numbers (EGFR-mediated survival of ICs) coupled to a correction mechanism inhibiting PGC proliferation. This feedback loop ensures that sufficient numbers of PGCs exist to fill all the stem-cell niches that form at the end of larval development. We propose that similar feedback mechanisms might be generally used for coordinated growth, regeneration and homeostasis.

Each ovary in the adult fruitfly *Drosophila melanogaster* is composed of 16–18 units called ovarioles. At the anterior of each ovariole, two or three germline stem cells (GSCs) interact with somatic cells that affect their establishment, maintenance and differentiation (Fig. 1a)³. These somatic niche cells develop across the larval gonad at the third larval instar (Fig. 1a) and are separated into ovarioles during early pupal development^{4–6}. The GSCs are derived from primordial germ cells (PGCs) that form in the early embryo. During the three larval instars, the entire gonad grows (compare Fig. 1b with Fig. 1c). The number of PGCs increases eightfold, from about 12 PGCs in each embryonic gonad⁷ to about 100 by the middle of third instar (ML3; Fig. 1d). PGCs double their numbers every 24 h during first (L1) and second (L2) instar, and division rates are slightly slower during the next 24 h (Fig. 1d). By ML3, sufficient PGCs exist to fill all the somatic GSC niches that form at that time.

When embryos contain very few PGCs (either when few PGCs are transplanted into embryos lacking germ cells, or in certain genetic backgrounds), they nonetheless develop into fully fertile females⁸. To test whether PGC proliferation might be regulated during larval growth, we counted PGC numbers in *germcell-less* (*gcl*) and *oskar* (*osk*) mutant embryos, in which fewer PGCs form during embryogenesis (A. Arkov and R.L., unpublished observations)⁹. In *gcl* mutants, in which only two PGCs on average (s.d. 1.2, $n = 31$) were incorporated into the embryonic gonad, an average of about 60 PGCs was reached by ML3 (Fig. 1d). This division rate (on average five divisions in three days) is higher than in the wild type (three divisions in three days). In *osk* mutants, in which 3 ± 1.1

(mean $\pm$ s.d., $n = 34$) PGCs were incorporated into the gonad, the division rate was faster than in the wild type until the end of L2 (EL2), but as PGCs approached wild-type numbers, their proliferation rate decreased markedly (Fig. 1d). Thus, PGC proliferation is regulated, and can increase or decrease to achieve wild-type numbers.

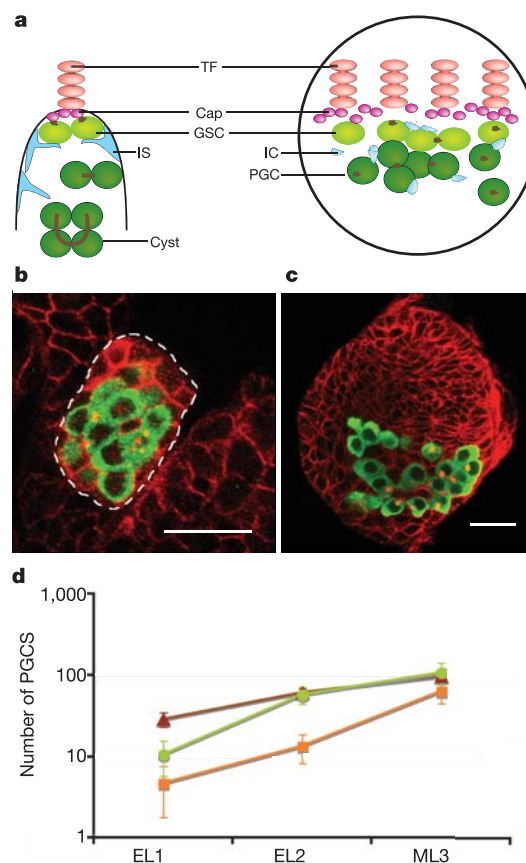


Figure 1 | The division of PGCs in the larval ovary is regulated. **a**, The adult germarium (left) and a late larval ovary (right). Terminal filament (TF), cap (Cap), inner sheath (IS) and intermingled cells (IC) are marked. **b**, **c**, Larval ovary at the end of embryogenesis (**b**) and in mid-third instar (**c**). Scale bars, 20 μ m. Anti-Vasa (green) labels PGCs; monoclonal antibody 1B1 (red) labels somatic cell membranes and the fusome, a round organelle within PGCs. **d**, PGC number in the wild type (maroon triangles) and in *gcl* (orange squares) and *osk* (green circles) mutants. Error bars represent s.d.. EL1, EL2 and ML3 are, respectively, end of first instar, end of second instar and middle of third instar.

¹Howard Hughes Medical Institute and Developmental Genetics Program, The Skirball Institute and Department of Cell Biology, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA.

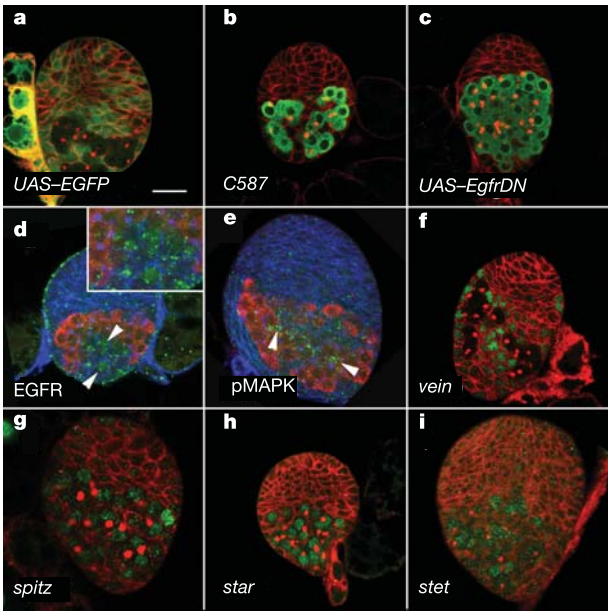


Figure 2 | Abrogation of somatic EGFR signalling causes PGC overproliferation. The monoclonal antibody 1B1 labels somatic cells red (a–c, f–i) or blue (d, e). a, *C587–Gal4* drives GFP expression within somatic cells of an ML3 gonad. PGCs lack GFP, and their round fusome is labelled by 1B1. *C587–Gal4* is also expressed in the fat body, contacting the gonad. b, c, EL2 gonads from *C587–Gal4* (b) or *C587–Gal4; UAS–EgfrDN*, where PGCs (anti-Vasa, green) overproliferate (c). d, e, Germ cells are labelled by anti-Vasa (red). Anti-EGFR (green) (d) and anti-pMAPK (green) (e) stain cells close to PGCs (arrowheads). Other somatic cells in the ovary may also express EGFR and pMAPK, but at much lower levels. f–i, Enhancer trap expression (anti-β-galactosidase, green) in *vein* (f), *spitz* (g), *star* (h) and *stet* (i). Scale bar, 20 μm.

To determine how PGC division rate is controlled, we used the Gal4/UAS system to express dominant-active or dominant-negative signalling pathway components in the larval gonad. For somatic expression we used the soma-specific *C587–Gal4* (Fig. 2a)⁶. For PGC-specific expression we used *nos–Gal4–VP16* (ref. 10). We observed a large increase in PGC numbers when the dominant-negative form of EGFR (*UAS–EgfrDN*) was expressed in somatic gonadal cells by *C587–Gal4* (Table 1 and Fig. 2b, c). Similarly, somatic overexpression of Ras85D.N17, a dominant-negative form of Ras85D, which acts downstream of EGFR, yielded an expansion of PGCs (Table 1). This indicates that EGFR might function in somatic gonadal cells to regulate PGC divisions.

To analyse how EGFR signalling affects PGC numbers, we determined the gonadal expression of different EGFR signalling components. Consistently with the somatic effect of *UAS–EgfrDN*, we detected EGFR and the phosphorylated form of mitogen-activated protein kinase (pMAPK) by antibody staining in somatic cells

adjacent to PGCs (Fig. 2d, e). Enhancer traps in both *vein* (Fig. 2f) and *argos* (not shown), which are known transcriptional targets of EGFR signalling, were also expressed in these cells¹¹. We found enhancer trap expression of the EGFR ligand Spitz in PGCs (Fig. 2g). Gurken, an additional EGFR ligand, which is expressed in oocytes, could not be detected in PGCs (not shown)¹². Ligand presentation requires Rhomboid family members (Rhomboid or Stet) and Star in the ligand-producing cells^{11,13}. We detected both *star* and *stet*, but not *rhomboid*, by enhancer trap expression in PGCs (Fig. 2h, i). These results indicate that PGCs might produce Spitz and activate EGFR signalling in neighbouring somatic cells.

Very little is known about the origin or function of the somatic cells adjacent to PGCs, which have been termed intermingled cells (IC)¹⁴. These cells express the *MA33* enhancer trap at L3 (Fig. 3a). ICs also express the protein Traffic Jam (TJ; Fig. 4c), although before L3 TJ expression is not limited to ICs (Supplementary Fig. 1)¹⁴. Our results demonstrate that the EGFR signalling pathway is activated in ICs. To determine its role in ICs we used a temperature-sensitive allelic combination of EGFR (*Egfr(ts)*) to overcome the earlier, embryonic, requirement for EGFR. PGCs overproliferated in *Egfr(ts)* larvae grown at the restrictive temperature, whereas the gonads of their wild-type siblings remained normal (Fig. 3b, c). Although ICs were readily observed in heterozygous gonads, fewer of them were present in *Egfr(ts)* gonads (compare Fig. 3b with Fig. 3c). A similar reduction in IC numbers was observed in *UAS–EgfrDN* gonads (Table 2). Antibodies against cleaved caspase 3, an apoptotic marker, revealed a significant increase ($P < 0.001$) in dying cells in *UAS–EgfrDN* gonads at EL2 (average 22, s.d. 8, $n = 13$) in comparison with those expressing *C587–Gal4* alone, (average 5, s.d. 4, $n = 15$) (Fig. 3d, e). Other aspects of somatic differentiation and morphogenesis seemed normal, because components of the niche such as terminal filament and cap cells differentiated normally in *Egfr(ts)* and *UAS–EgfrDN* flies (not shown). Thus, EGFR signalling is required for IC survival. A similar role for EGFR signalling in cell survival has been described in the *Drosophila* nervous system, in which neuronal cells secrete Spitz and protect midline glial cells from death¹⁵. To determine whether IC death resulted directly from the abrogation of EGFR signalling, or indirectly from PGC overproliferation, we mis-expressed *UAS–CycD*, *UAS–Cdk4* in PGCs¹⁶. Under these conditions, PGCs overproliferated extensively, without loss of ICs (Fig. 3f). This indicates that PGC overproliferation in *UAS–EgfrDN* and in *Egfr(ts)* might have been the effect rather than the cause of IC death.

To ask more directly whether PGC production of Spitz is required for IC survival, we reduced Spitz production by RNA interference (*UAS–spiRNAi*) in either somatic gonadal cells or PGCs. Expression of *UAS–spiRNAi* in PGCs resulted in a significant increase in PGC numbers, whereas somatic expression had no effect (Table 1). As could be expected, reducing Spitz production in PGCs resulted in reduced IC numbers (Table 2). We then examined EGFR signalling in gonads lacking PGCs altogether. At L1 and L2, pMAPK was absent (Fig. 4a). By L3, pMAPK was detected in a subpopulation of

Table 1 | Effects of mis-expression of EGFR signalling components on PGC numbers

Genotype	No. of PGCs at EL2 (s.d., n)	P	No. of PGCs at ML3 (s.d., n)	P
<i>C587-GAL4</i>	61.25 (7, 32)	NA	97 (13, 23)	NA
<i>C587-GAL4/+; UAS-EgfrDN/+</i>	113.6 (25, 24)	<0.0001	204 (8, 18)	<0.0001
<i>C587-GAL4/UAS-Ras85D.N17</i>	86.9 (13, 12)	<0.0001	146 (22, 23)	<0.0001
<i>nos-Gal4-VP16/UAS-spiRNAi</i>	ND		122 (25, 40)	<0.0001
<i>C587-GAL4/UAS-spiRNAi</i>	ND		102 (25, 35)	0.27
<i>C587-GAL4/+; UAS-EgfrCA/+</i>	45.23 (7, 14)	<0.0001	57 (12, 23)	<0.0001
<i>C587-GAL4/+; UAS-phl.gof/+</i>	47.2 (6, 11)	<0.0001	60 (13, 13)	<0.0001
<i>C587-GAL4/+; UAS-Ras85D.GV12/+</i>	54.6 (9, 11)	0.02	52 (11, 17)	<0.0001
<i>nos-GAL4-VP16/UAS-sSpi</i>	34.9 (9, 17)	<0.0001	51 (22, 35)	<0.0001

Cell numbers, s.d., sample size (n) and P values of Student's t-test, which compare the wild type (*C587-GAL4*) with the experimental group, are shown. The normality of each data set was confirmed by Kolmogorov–Smirnov tests. ND, not determined; NA, not applicable.

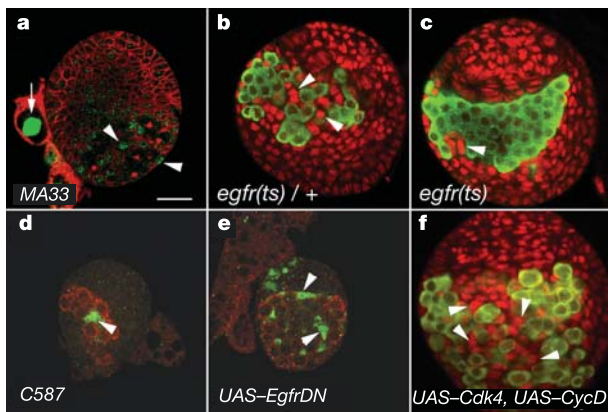


Figure 3 | Death of IC causes PGC overproliferation. **a**, The enhancer trap MA33 (anti- β -galactosidase, green) labels IC in ML3 gonads (arrowheads). It also expresses strongly in fat body nuclei (arrow), 1B1 antibody, red. **b, c**, Very few ICs (anti-Zfh-1, red) are observed in *egfr(ts)* homozygous ML3 gonads (**c**), in comparison with heterozygotes (**b**). Anti-Vasa, green. **d, e**, Extensive death of ICs is marked by anti-cleaved caspase 3 (green) in *UAS-EgfrDN* EL2 gonads (**e**), but not in the wild-type (**d**). PGCs are labelled by anti-Vasa (red). **f**, ICs can be observed when PGCs overproliferate in *UAS-CycD*, *UAS-Cdk4;nos-GAL4* ML3 gonads. Anti-Zfh1 labels somatic nuclei (red). Anti-Vasa (green). Scale bar, 20 μ m.

migrating cells but not where ICs are located (not shown). In *gcl* mutants containing PGCs, the strongest pMAPK staining was observed in somatic cells contacting PGCs (Fig. 4b). In *gcl* gonads lacking PGCs, ICs could be detected by TJ expression, but in greatly reduced numbers (Table 2; compare Fig. 4c with Fig. 4d). This is not due to a general reduction in somatic cell numbers, because similar numbers of terminal filaments form in *gcl* and in wild-type gonads (not shown). The reduction in IC number resembles the disappearance of inner-sheath cells from adult germaria lacking germ cells^{17,18}. MA33 could not be detected in *gcl* gonads (not shown). The difference could be due to weaker staining of MA33 than that of TJ, or because MA33-positive cells represent a subpopulation of ICs that disappears in gonads lacking PGCs. Taken together, our results indicate that PGCs produce Spitz and activate EGFR signalling in ICs, which is necessary for their survival. In return, ICs inhibit PGC proliferation (Fig. 4f). We suggest that this feedback mechanism allows the gonad to monitor and correct PGC numbers during larval growth. When very few PGCs form, Spitz production is low, leading to reduced IC numbers. This, in turn, leads to increased PGC proliferation. Compensation of PGC numbers by the end of larval development ensures that sufficient PGCs are present to occupy the adult niches.

To test whether EGFR signalling in ICs has an additional role to that of promoting survival, we increased EGFR signalling by mis-expressing constitutively active forms of EGFR signalling components (*UAS-EgfrCA*, *UAS-Ras85D.G12V* or *UAS-phl.gof*) in the soma, or mis-expressing the secreted form of Spitz in PGCs

Table 2 | Effects of mis-expression of EGFR signalling components on IC numbers

Genotype	No. of ICs at ML3 (s.d., n)	P
C587-GAL4	344 (35, 23)	NA
C587-GAL4/+; <i>UAS-EgfrDN</i> /+	89 (17, 8)	<0.0001
<i>gcl</i>	142 (40, 8)	<0.0001
C587-GAL4/+; <i>UAS-EgfrCA</i> /+	352 (30, 13)	0.314
<i>nos-GAL4</i> -VP16/ <i>UAS-sSpi</i>	325 (33, 11)	0.15
<i>nos-GAL4</i> -VP16/ <i>UAS-spiRNAi</i>	299 (27, 23)	<0.0001

Cell numbers, s.d., sample size (n) and P values of Student's t-test, which compare the wild type (C587-GAL4) with the experimental group, are shown. The normality of each data set was confirmed by Kolmogorov-Smirnov tests. NA, not applicable.

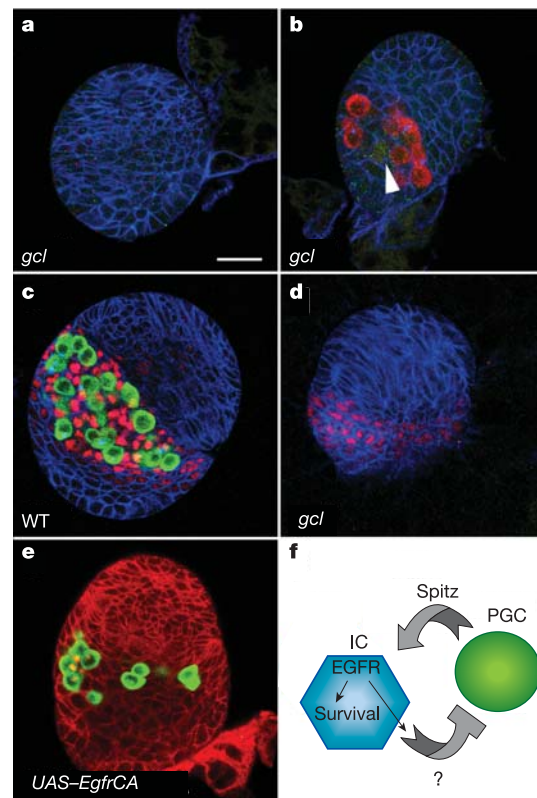


Figure 4 | PGCs determine intermingled cell number. **a–d**, The monoclonal antibody 1B1 labels somatic cells (blue). **a, b**, *gcl* mutants at EL2. The strongest pMAPK (green, arrowhead) signal can be detected in ICs only when PGCs (anti-Vasa, red) are incorporated in the gonad. **c, d**, Germ cells are marked with anti-Vasa (green). More ICs (anti-TJ, red) can be observed in wild-type (WT) gonads (**c**) than in gonads lacking a germ line (**d**). **e**, Fewer PGCs (anti-Vasa, green) are present in *UAS-EgfrCA* gonads than in the wild type, leading to niches without PGCs (compare **c** with **e**). Monoclonal antibody 1B1 labels somatic cells (red). **f**, A model describing a feedback loop between PGCs and ICs; see the text for details. Scale bar, 20 μ m.

(*UAS-sSpi*). PGC numbers were significantly reduced in these cases (Table 1 and Fig. 4e). Interestingly, IC numbers remained unchanged (Table 2). The restriction of TJ expression to ICs was also unchanged (Supplementary Fig. 1). Because increasing EGFR signalling resulted in decreased PGC numbers, without an apparent effect on gonad morphogenesis or IC numbers, we suggest that EGFR signalling in ICs might be directly required for the inhibition of PGC proliferation (Fig. 4f).

Soma–germline interactions through EGFR signalling are a recurring motif in *Drosophila*. In females they serve to pattern the eggshell and localize the oocyte nucleus¹⁹. In males they serve to restrict GSC proliferation and promote differentiation^{20,21}. The *rhomboid* homologue *stet* is required in both males and females for GSC differentiation and for proper connections between somatic cells and germ cells¹³. The signals originating in the somatic cells and perceived by germ cells remain unknown. In this study we show that EGFR has a central role in a feedback loop coordinating IC survival and PGC proliferation (Fig. 4f). The properties of this loop make it ideal for regulating homeostasis and for coordinating the growth of different cell populations in any organ. In the liver, for example, several cell types proliferate after injury. It has been suggested that hepatocytes provide the mitogenic stimuli for other liver cells, such as Kupffer cells, hepatic stellate cells and biliary ductular cells. The production of transforming growth factor β by hepatic stellate cells may, in turn, limit hepatic growth². Similar

feedback loops may apply in other cases during normal development or after injury.

METHODS

Fly stocks. *UAS-SpiRNAi* flies were a gift from G. Dietzl and B. Dickson. *gcl* mutants were a heterozygous combination of a strong *gcl* allele (*A43-33*; A. Arkov and R.L., unpublished) and the deficiency *Df(2R)H3E1*. *osk* mutants were larvae produced by mothers *trans*-heterozygous for *osk(166)* and *osk(GM52)* (ref. 9). *egfr(ts)* was an allelic combination of *egfr(tsla)* and *egfr(f24)* (FlyBase). The somatic driver *C587-GAL4* (ref. 6), the germline driver *nos-GAL4-VP16* (ref. 10) and the enhancer trap *MA33* (ref. 22) were as described. A description of other enhancer trap lines and the various UAS constructs can be found in FlyBase.

Staging of larvae and antibody staining. Staging was as described in ref. 23.

Antibodies. Antibodies were used in the following concentrations: 1B1 mouse monoclonal (dilution 1:20) was from the Developmental Studies Hybridoma Bank; chicken anti-Vasa 1:10,000 was from the Lehmann laboratory; rabbit anti-Vasa 1:5,000 from the Lehmann laboratory; rat anti-Zfh1 1:700 from the Lehmann laboratory; rabbit anti-EGFR 1:1,000 was a gift from A. Rodrigues and K. Moses²⁴; guinea-pig anti-Traffic Jam 1:5,000 was a gift from D. Godt¹⁴. Rabbit anti- β -galactosidase 1:20,000 (Cappel), rabbit anti-phospho-histone H3 (Upstate Biotechnology), rabbit anti-cleaved caspase-3 (Cell Signaling) and rabbit anti-pMAPK (Cell Signaling) were used in accordance with the manufacturers' protocols.

Secondary antibodies were purchased from Jackson ImmunoResearch Laboratories, and were used at a dilution of 1:500.

Temperature shifts for *egfr(ts)*. Heterozygous males and females were allowed to lay for 2 h. The bottles were kept at 18 °C for two days, to successfully complete embryogenesis, then shifted to 29 °C. Homozygous mutant larvae and their heterozygous wild-type siblings were dissected at ML3 or late L3.

Immunofluorescence. Fixation and immunostaining of gonads were performed in accordance with standard protocols. Imaging was performed on a Leica DM RBE confocal microscope using the Leica TCS NT program or on a Zeiss LSM 510 META using the LSM software.

Received 4 May; accepted 10 July 2006.

Published online 27 August 2006.

- Freitas, A. A. & Rocha, B. Population biology of lymphocytes: the flight for survival. *Annu. Rev. Immunol.* **18**, 83–111 (2000).
- Michalopoulos, G. K. & DeFrances, M. C. Liver regeneration. *Science* **276**, 60–66 (1997).
- Gilboa, L. & Lehmann, R. How different is Venus from Mars? The genetics of germ-line stem cells in *Drosophila* females and males. *Development* **131**, 4895–4905 (2004).
- Godt, D. & Laski, F. A. Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of *bric a brac*. *Development* **121**, 173–187 (1995).
- Song, X., Zhu, C. H., Doan, C. & Xie, T. Germline stem cells anchored by adherens junctions in the *Drosophila* ovary niches. *Science* **296**, 1855–1857 (2002).
- Zhu, C. H. & Xie, T. Clonal expansion of ovarian germline stem cells during niche formation in *Drosophila*. *Development* **130**, 2579–2588 (2003).
- Poirié, M., Niederer, E. & Steinmann-Zwicky, M. A sex-specific number of germ cells in embryonic gonads of *Drosophila*. *Development* **121**, 1867–1873 (1995).
- Robertson, S. E., Dockendorff, T. C., Leatherman, J. L., Faulkner, D. L. & Jongens, T. A. *germ cell-less* is required only during the establishment of the germ cell lineage of *Drosophila* and has activities which are dependent and independent of its localization to the nuclear envelope. *Dev. Biol.* **215**, 288–297 (1999).
- Yohn, C. B., Pusateri, L., Barbosa, V. & Lehmann, R. *I(3)malignant brain tumor* and three novel genes are required for *Drosophila* germ-cell formation. *Genetics* **165**, 1889–1900 (2003).
- Van Doren, M., Williamson, A. & Lehmann, R. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Curr. Biol.* **8**, 243–246 (1998).
- Shilo, B. Z. Signaling by the *Drosophila* epidermal growth factor receptor pathway during development. *Exp. Cell Res.* **284**, 140–149 (2003).
- Nilson, L. A. & Schupbach, T. EGF receptor signaling in *Drosophila* oogenesis. *Curr. Top. Dev. Biol.* **44**, 203–243 (1999).
- Schulz, C., Wood, C. G., Jones, D. L., Tazuke, S. I. & Fuller, M. T. Signaling from germ cells mediated by the rhomboid homolog *stet* organizes encapsulation by somatic support cells. *Development* **129**, 4523–4534 (2002).
- Li, M. A., Alls, J. D., Avancini, R. M., Koo, K. & Godt, D. The large Maf factor Traffic Jam controls gonad morphogenesis in *Drosophila*. *Nature Cell Biol.* **5**, 994–1000 (2003).
- Bergmann, A., Tugentman, M., Shilo, B. Z. & Steller, H. Regulation of cell number by MAPK-dependent control of apoptosis: a mechanism for trophic survival signaling. *Dev. Cell* **2**, 159–170 (2002).
- Datar, S. A., Jacobs, H. W., de la Cruz, A. F., Lehner, C. F. & Edgar, B. A. The *Drosophila* cyclin D-Cdk4 complex promotes cellular growth. *EMBO J.* **19**, 4543–4554 (2000).
- Kai, T. & Spradling, A. An empty *Drosophila* stem cell niche reactivates the proliferation of ectopic cells. *Proc. Natl Acad. Sci. USA* **100**, 4633–4638 (2003).
- Margolis, J. & Spradling, A. Identification and behavior of epithelial stem cells in the *Drosophila* ovary. *Development* **121**, 3797–3807 (1995).
- Van Buskirk, C. & Schupbach, T. Versatility in signalling: multiple responses to EGF receptor activation during *Drosophila* oogenesis. *Trends Cell Biol.* **9**, 1–4 (1999).
- Tran, J., Brenner, T. J. & DiNardo, S. Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis. *Nature* **407**, 754–757 (2000).
- Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**, 750–754 (2000).
- Roth, S., Neumann-Silberberg, F. S., Barcelo, G. & Schupbach, T. *cornichon* and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in *Drosophila*. *Cell* **81**, 967–978 (1995).
- Gilboa, L. & Lehmann, R. Repression of primordial germ cell differentiation parallels germ line stem cell maintenance. *Curr. Biol.* **14**, 981–986 (2004).
- Rodrigues, A. B., Werner, E. & Moses, K. Genetic and biochemical analysis of the role of *Egfr* in the morphogenetic furrow of the developing *Drosophila* eye. *Development* **132**, 4697–4707 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Treisman for her support with both materials and ideas; G. Dietzl and B. Dickson for sharing *UAS-SpiRNAi* before publication; B. Edgar and L. Johnston for materials; J. Lafaille for discussions; S. Burden and J. Morris for critical reading of the manuscript; and C. Navarro, D. Siekhaus and all members of the Lehmann laboratory for comments on the manuscript. The Bloomington Stock Center provided reagents. L.G. is supported by a fellowship from the Helen and Martin Kimmel Center for Stem Cell Biology. R.L. is a Howard Hughes Medical Institute investigator.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.L. (lehmann@saturn.med.nyu.edu).

Sperm chromatin proteomics identifies evolutionarily conserved fertility factors

Diana S. Chu¹, Hongbin Liu^{2,3}, Paola Nix^{4,5}, Tammy F. Wu¹, Edward J. Ralston^{4,5}, John R. Yates III² & Barbara J. Meyer^{4,5}

Male infertility is a long-standing enigma of significant medical concern. The integrity of sperm chromatin is a clinical indicator of male fertility and *in vitro* fertilization potential¹: chromosome aneuploidy and DNA decondensation or damage are correlated with reproductive failure. Identifying conserved proteins important for sperm chromatin structure and packaging can reveal universal causes of infertility. Here we combine proteomics, cytology and functional analysis in *Caenorhabditis elegans* to identify spermatogenic chromatin-associated proteins that are important for fertility. Our strategy employed multiple steps: purification of chromatin from comparable meiotic cell types, namely those undergoing spermatogenesis or oogenesis; proteomic analysis by multidimensional protein identification technology (MudPIT) of factors that co-purify with chromatin; prioritization of sperm proteins based on abundance; and subtraction of common proteins to eliminate general chromatin and meiotic factors. Our approach reduced 1,099 proteins co-purified with spermatogenic chromatin, currently the most extensive catalogue, to 132 proteins for functional analysis. Reduction of gene function through RNA interference coupled with protein localization studies revealed conserved spermatogenesis-specific proteins vital for DNA compaction, chromosome segregation, and fertility. Unexpected roles in spermatogenesis were also detected for factors involved in other processes. Our strategy to find fertility factors conserved from *C. elegans* to mammals achieved its goal: of mouse gene knockouts corresponding to nematode

proteins, 37% (7/19) cause male sterility. Our list therefore provides significant opportunity to identify causes of male infertility and targets for male contraceptives.

Basic features of DNA compaction and partitioning differ during spermatogenesis and oogenesis², but sperm-specific processes are conserved, making *C. elegans* appropriate for identifying and functionally validating male fertility factors. Whereas spermatocytes complete meiotic divisions rapidly to produce four haploid spermatids, oocytes complete meiosis after fertilization, creating one haploid cell and two polar bodies (Fig. 1a). In most species sperm DNA is packaged uniquely from that of oocytes using small basic structural proteins called sperm nuclear basic proteins (SNBPs). SNBPs in *C. elegans* are yet to be identified, as are factors in any organism that mediate somatic histone displacement, incorporate SNBPs or implement other sperm-specific processes such as transcriptional silencing.

We developed a strategy to identify abundant proteins that co-purify with spermatogenic chromatin and analysed their functions in fertility (Fig. 1b). Chromatin was isolated from populations of spermatogenic or oogenic germ nuclei (Methods). Proteins co-purified with chromatin were identified by using a combination of MudPIT (mass spectrometric analysis to acquire tandem mass spectra³) coupled with the SEQUEST algorithm to match spectra to predicted peptides⁴ and DTASelect to filter and assemble peptides into corresponding *C. elegans* proteins⁵. In all, 1,099 spermatogenic and 812 oogenic proteins were identified (Supplementary Tables 1–4).

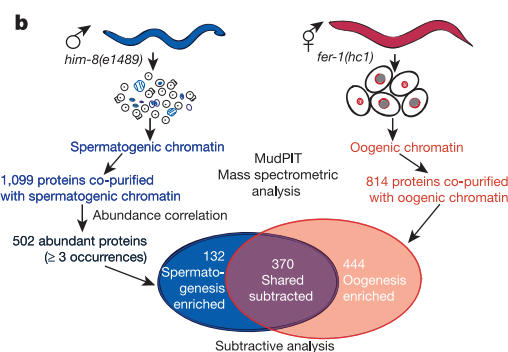
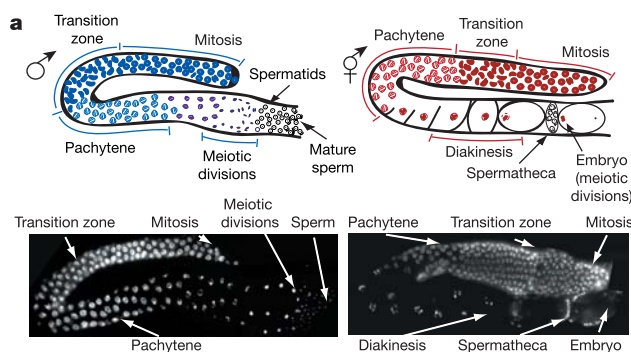


Figure 1 | Abundance-correlated subtractive proteomic strategy to identify spermatogenesis-enriched proteins. **a**, Gametogenesis in *C. elegans*. Upper panels illustrate the progression of germ-cell nuclei in a wild-type male gonad (left) and in one arm of a wild-type bilaterally symmetric hermaphrodite gonad (right). Lower panels show corresponding DAPI-stained nuclei from dissected and fixed gonads. Visually distinct stages of gametogenesis are labelled. **b**, Proteomic strategy. Spermatogenic chromatin was isolated from germ nuclei of adult XO males derived from the

him-8(e1489) X chromosome non-disjunction mutant. Oogenic chromatin was isolated from germ cells of *fer-1(hc1)* hermaphrodites. *fer-1(hc1)* mutants produce defective sperm, causing XX animals to be functional females in which some of the oocytes mature, are ovulated unfertilized, and undergo endoreduplication. Spermatogenic and oogenic chromatin was subjected to MudPIT for identification of associated proteins. Subtractive analysis of chromatin proteins identified 132 abundant spermatogenesis-enriched proteins.

¹Department of Biology, 1600 Holloway Avenue, San Francisco State University, San Francisco, California 94132, USA. ²The Scripps Research Institute, Department of Cell Biology, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ³Agilent Technologies, Inc., 2850 Centerville Road, Wilmington, Delaware 19808, USA. ⁴Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA. ⁵Howard Hughes Medical Institute, Berkeley, California 94720, USA.

Spermatogenic proteins were prioritized on the basis of abundance measurements derived from MudPIT, which was an advance over previous studies^{6,7}. By choosing only abundant spermatogenic proteins that occurred in three or more preparations, we reduced the number of candidate proteins from 1,099 to 502 (Methods). The 597 eliminated proteins constitute only 12% of the relative mass of total spermatogenic proteins (Supplementary Table 5a).

A further reduction in the number of abundant proteins under consideration was achieved by subtracting factors common to spermatogenic and oogenic chromatin samples, eliminating 74% (370 of 502) of spermatogenic proteins (Supplementary Methods and Supplementary Table 5b, c). We expected the remaining proteins to include spermatogenesis-specific proteins and spermatogenesis-enriched proteins expressed in low abundance in oocytes. Thus, from 1,099 initial candidates that co-purify with chromatin, our strategy focused functional analysis on 132 abundant spermatogenic proteins (Supplementary Table 1).

The subcellular localization of 11 candidate proteins demonstrates that spermatogenesis-specific and highly spermatogenesis-enriched chromatin factors were correctly identified. By immunolocalization, eight proteins, representing four different families (GSP-3, GSP-4 (Glc-seven phosphatase), SMZ-1, SMZ-2 (sperm meiosis PDZ), HTAS-1 (histone two A sperm) and SPCH-1, SPCH-2, SPCH-3 (sperm chromatin enriched)), associate specifically with spermatogenic meiotic chromosomes and mature sperm chromatin (Fig. 2b–e and Table 1). These proteins resemble the single known *C. elegans* sperm-specific chromatin-associated protein SPE-11 (spermatogenesis defective), also identified by our proteomic analysis. SPE-11 is a paternal factor supplied to eggs for subsequent use in embryonic development⁸ (Fig. 2a and Table 1). Three other proteins (TOP-1 (topoisomerase), GLH-2 (germ line helicase) and HCP-4^{CENP-C} (holocentric chromosome-binding protein) known to function in somatic cells^{9–11} also associate with sperm chromatin, indicating possible roles in spermatogenesis (Fig. 2f, h, Table 1 and Supplementary Fig. 3). In sum, all 11 tested candidates revealed previously undetected associations with spermatogenic chromatin.

Functional analysis of the 132 proteins further validated our approach. The importance of each protein for fertility was evaluated with the use of RNA interference (RNAi) against cognate genes to reduce gene products. The sterility of treated animals was assayed by brood counts, and chromosome or gonad abnormalities were assessed by cytology (Methods). Although genes required for spermatogenesis are notoriously insensitive to RNAi (S. Ward and S. L'Hernault, personal communication), sterility or embryonic lethality was found for 50 of 132 genes (Supplementary Table 6) and germline cytological defects for 20 of these 50 genes (Table 1 and Supplementary Table 7). RNAi treatment of 18 of 20 genes disrupted aspects of male fertility, including sperm meiotic chromosome segregation and male germline morphology (Table 1, Supplementary Table 7 and Supplementary Figs 4 and 5). Thus, RNAi analysis showed spermatogenesis-enriched proteins to be important for fertility and embryonic development.

Proteomic identification of factors co-purified with sperm chromatin coupled with injection RNAi and quantitative screening resulted in a higher rate for identifying reproductive factors than that for previous studies. Our rate (38%; 50 of 132) is tenfold that (less than 3%) for RNAi-treated genes detected by microarray analysis for enriched transcript levels during spermatogenesis versus oogenesis¹². Proteomic analysis identified factors overlapping with (42) and differing from (90) genes transcriptionally enhanced during spermatogenesis. Sterility and embryonic lethality caused by RNAi were observed for 41% of non-overlapping and 31% of overlapping factors. Our rate is also threefold that (10%) for identifying fertility genes through *C. elegans* genome-wide RNAi screens^{13,14}.

Many *C. elegans* homologues of mammalian fertility factors were identified: 29 of 132 *C. elegans* proteins (representing 24 protein families) correspond to 19 genetically disrupted mouse homologues

(Table 1 and Supplementary Table 8); 37% (7 of 19) of mouse knockouts (corresponding to 14 worm homologues in 7 families) caused male infertility. Half (7 of 14) of these worm homologues function in spermatogenesis based on RNAi phenotype (GSP-3 and GSP-4) or have specific association with spermatogenic chromatin (HTAS-1, SPCH-1, SPCH-2, SPCH-3 and GLH-2). Our subtractive proteomics approach is therefore valid for finding evolutionarily conserved, spermatogenesis-enriched proteins essential for fertility.

Analysis of these diverse proteins in *C. elegans* revealed essential functions in spermatogenesis, including roles in meiotic chromosome segregation and chromosome condensation and architecture. Analysed proteins were divided into three categories on the basis of protein localization and RNAi phenotypes: category I proteins have spermatogenesis-specific localization, category II proteins have newly discovered roles in spermatogenesis and previously described roles in other cellular processes (Table 1), and category III proteins have other roles in fertility (Supplementary Table 7).

Category I proteins GSP-3 and GSP-4 are nearly identical homologues of PP1- γ , a glc7/PP1 phosphatase required for mouse male meiosis and spermiogenesis¹⁵. GSP-3 and GSP-4 encase chromosomes during meiotic divisions and in mature sperm (Fig. 2b). RNAi against either caused variably penetrant male sterility as a result of chromosome segregation defects, including chromosome bridges and incompletely separated chromosomes (Supplementary Fig. 4a–d). Previous *gsp-3* and *gsp-4* RNAi experiments showed hermaphrodite sterility^{16,17}. Fertility in *gsp-4*(RNAi) (Supplementary Table 6) and *gsp-3*(RNAi)¹⁷ hermaphrodites was restored by mating with wild-type males, indicating defective spermatogenesis. GSP-3 and GSP-4 may act as the sperm-specific counterparts to the other *C. elegans* Glc7/PP1 phosphatases, GSP-1 and GSP-2, which antagonize the Aurora B kinase-mediated release of chromosome cohesion during mitosis and oocyte meiosis¹⁸. Hence, studies in worms and mice suggest that GSP family members act directly on chromatin to achieve proper chromosome segregation and fertility in various species.

Also identified were spermatogenesis-specific proteins that lack homology to known fertility factors, yet are essential for sperm meiotic chromosome segregation and male fertility. SMZ-1 and SMZ-2 are 89% identical, contain PDZ (protein–protein interaction) domains, and represent a novel protein family vital for spermatogenic chromosome segregation. These proteins localize to male germline nuclei during and after meiosis and concentrate around sperm chromatin in 30–50% of spermatids (Fig. 2c). *smz-1*(RNAi) or *smz-2*(RNAi) spermatocytes failed to progress through meiotic divisions. Defective meiotic chromosomes did not congress to the metaphase plate or segregate (Supplementary Fig. 4e). Thus, our approach identified highly conserved and *C. elegans* sperm-specific proteins that are important for male meiosis and fertility.

Many organisms use SNBPs such as histone variants and protamines to establish unique sperm chromosome structure. In mammals, histone H2A variants such as H2AX substitute for H2A during spermatogenesis and are required for male fertility¹⁹. We identified the first *C. elegans* SNBPs, including the histone H2A variant HTAS-1, which has 51% identity to canonical histones. It localizes with meiotic chromosomes and persists on mature sperm chromatin (Fig. 2e). Histone variants are therefore part of the unique constitution of sperm chromatin in worms and mammals.

Proteomic analysis further identified SPCH-1, SPCH-2 and SPCH-3, three small, abundant, highly basic proteins, which resemble invertebrate protamines such as the surf clam SNBP, PL-I (BLAST expectation score (*E* value) 10^{-23})²⁰. As expected for SNBPs, SPCH proteins are spermatogenesis-specific, localize to meiotic DNA and encase mature sperm chromatin (Fig. 2d). They are also homologous to mouse HANP/HIT2 (ref. 21) (Table 1), which localizes to mouse sperm chromatin, is required for the nuclear localization of protamines and is important for male fertility²². Although the rapid evolution of male reproductive genes²³ has made *C. elegans* SNBPs difficult to identify, proteomic analysis pinpointed SNBPs

with similar features in vertebrates and invertebrates.

Unknown roles in DNA compaction and chromosome segregation during spermatogenesis were revealed for proteins with functions in other cell types (category II; Table 1). The topoisomerase I homologue TOP-1 is a nucleolar protein in somatic and germ cells⁹, but unlike the subtracted nucleolar protein FIB-1 (Fig. 2g), TOP-1 surrounds mature sperm chromatin (Fig. 2f) and functions during spermatogenesis. Abnormally large sperm nuclei and aberrant progression through male meiosis were caused by *top-1*(RNAi), possibly indicating either defective DNA condensation or abnormal DNA content (Supplementary Fig. 5a, b).

Because topoisomerases alter DNA twist and writhe, TOP-1 may resolve *C. elegans* chromosomes during meiotic divisions, condense DNA during spermatogenesis or decondense DNA after fertilization, which are functions related to those of yeast topoisomerase I in mitosis²⁴. Alternatively, because human topoisomerase I is a kinase that phosphorylates SR proteins (key regulators of RNA processing events such as splicing)²⁵, TOP-1 could control RNA regulation during sperm formation. In fact, *top-1*(RNAi) defects resembled those caused by RNAi of *rsp-6*, which encodes an SR protein (Table 1 and Supplementary Fig. 5b, c). Other sperm-enriched SR proteins, RSP-1 (three occurrences) and RSP-2 (one occurrence), function together during spermatogenesis: double-mutant males are sterile²⁶.

Of identified factors, 23% (30 of 132) show homology to trans-

lation or splicing factors, many belonging to large families. Our subtractive proteomic analysis differentiated between family members to identify those with spermatogenesis-specific chromosome association or function. For example, unlike the subtracted P-granule RNA helicase GLH-1 (Fig. 2i), the GLH-2 family member surrounds sperm meiotic chromosomes (Fig. 2h). Although disruption of *glh-2* had no effect, disruption of *vbh-1* (vasa and belle-like helicase; two occurrences) caused male sterility (R. Navarro, personal communication). Similarly, disruption of the homologous RNA helicase Mvh (murine VASA) causes male infertility²⁷.

Our study provides clues to the underlying causes of mammalian infertility by identifying and analysing spermatogenesis-enriched chromatin proteins in the nematode *C. elegans*. Cross-species comparison of sperm proteomes, characterized by approaches similar to ours, will elucidate the molecular basis for sperm evolution and male fertility.

Our list of proteins represents a resource for identifying causes of human infertility, thereby providing candidates for any disease locus residing in a multigenic region correlated with a fertility defect. One illustration is azoospermia, the complete lack of sperm²⁸. *DBY*, which encodes a VASA RNA helicase, falls within a Y-chromosome region correlated with this defect, and is further implicated in spermatogenesis by our analysis.

Of the 132 proteins detected by our strategy, 70 (53%) have human

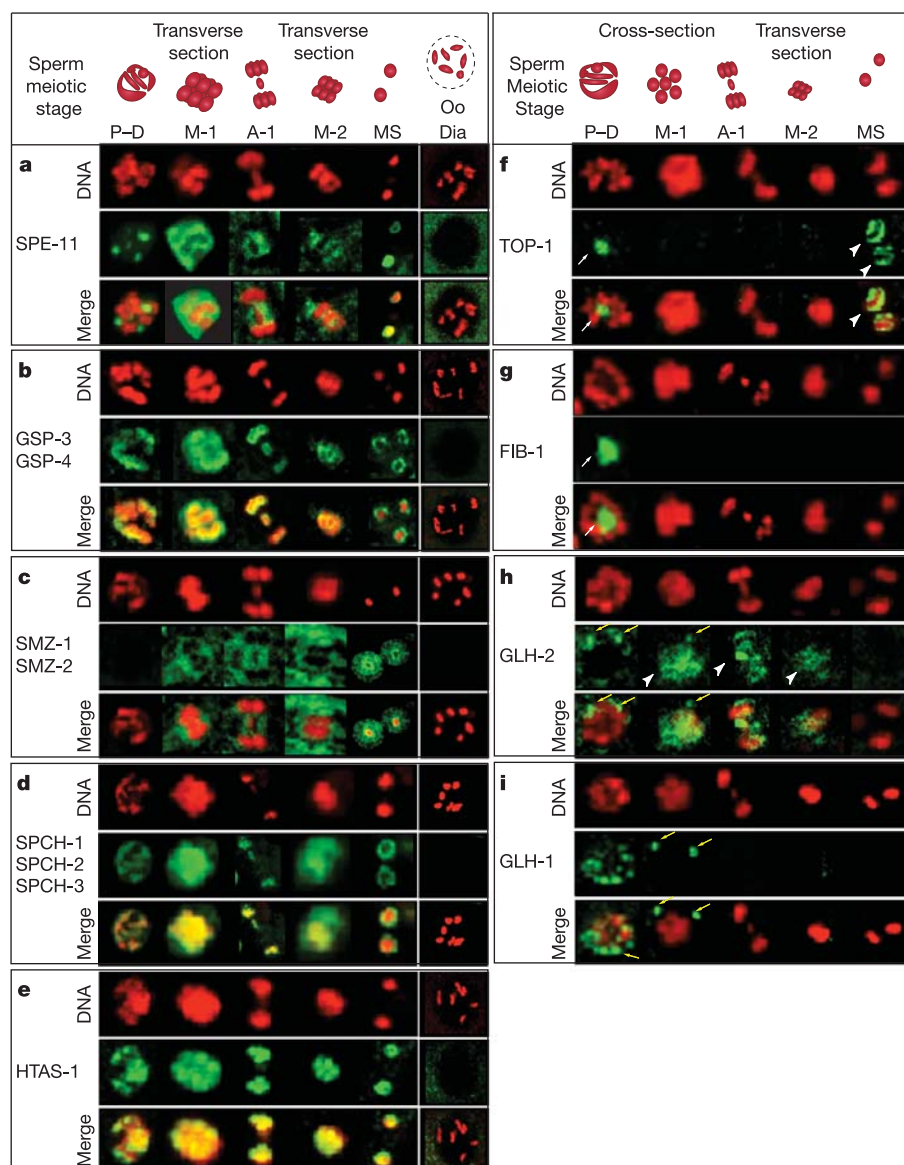


Figure 2 | Identification of spermatogenic chromatin-associated proteins by proteomic analysis. Immunolocalization of spermatogenic proteins show chromatin association in mature sperm and during meiotic divisions. Sperm meiotic stage: P-D, pachytene to diakinesis transition; M-1, metaphase I; A-1, anaphase I; M-2, metaphase II; MS, mature sperm. Oogenesis: Oo Dia, oocyte diakinesis. DNA is shown in red, antibody staining in green. **a-e**, Spermatogenesis-specific chromatin-associated proteins (proteins were not detected during any stage of oogenesis): **a**, SPE-11; **b**, GSP-3 and GSP-4; **c**, SMZ-1 and SMZ-2; **d**, SPCH-1, SPCH-2 and SPCH-3; **e**, HTAS-1. **f, g**, Spermatogenic localization of nucleolar proteins: **f**, TOP-1; **g**, FIB-1 (*C. elegans* fibrillar). TOP-1 surrounds mature sperm chromatin (white arrowheads) but FIB-1 does not (white arrows). Nucleolar localization (white arrows) is ubiquitous in spermatogenic and oogenic germ lines (data not shown, and ref. 9). **h, i**, Spermatogenic localization of P-granule components: **h**, GLH-1; **i**, GLH-2. GLH-2 associates with sperm meiotic chromatin (white arrowheads) but GLH-1 does not. Both localize to perinuclear P-granules (yellow arrows) during spermatogenesis and oogenesis (data not shown, and ref. 10).

Table 1 | Abundant *C. elegans* spermatogenesis-enriched chromatin proteins with roles in fertility

Gene ID	Gene name	Descriptor	Caenorhabditis elegans					Mammalian	
			Localization	RNAi cytological defects		Phenotype	Sum: new evidence for fertility function	Known homologue	Fertility link
				Male	Hermaphrodite				
Category I									
F48C1.7	spe-11	Paternal factor required for development ⁸	High levels on sperm DNA during and after meiotic divisions	None detected	Embryos do not form hard eggshells, Emb (mutant)	Male Ste (mutant)/(RNAi resistant)			
W09C3.6*	gsp-3	Glc7/PP1 phosphatase	High levels in sperm, strong around meiotic and mature sperm DNA	Defective sperm chromosome segregation	Embryos do not form hard eggshells, Emb	High male Ste	Spermatogenesis-specific localization/meiotic sperm chromosome segregation defects	M, Ppp1cc (1 × 10 ⁻¹⁰²) H, PPP1C (1 × 10 ⁻⁹⁵)	M, Ppp1cc KO male infertile
T03F1.5*	gsp-4	Glc7/PP1 phosphatase	High levels in sperm, strong around meiotic and mature sperm DNA	Defective sperm chromosome segregation	Embryos do not form hard eggshells, Emb	High male Ste	Spermatogenesis-specific localization/meiotic sperm chromosome segregation defects	M, Ppp1cc (1 × 10 ⁻¹⁰²) H, PPP1C (1 × 10 ⁻⁹⁵)	M, Ppp1cc KO male infertile
C25G4.6†	smz-1	PDZ protein-protein interaction domain	High levels in sperm, strong around 50% of mature sperm DNA	Defective sperm chromosome segregation	Embryos do not form hard eggshells, Emb	High male Ste	Spermatogenesis-specific localization/meiotic sperm chromosome segregation defects		
T21G5.4†	smz-2	PDZ protein-protein interaction domain	High levels in sperm, strong around 50% of mature sperm DNA	Defective sperm chromosome segregation	Embryos do not form hard eggshells, Emb	High male Ste	Spermatogenesis-specific localization/meiotic sperm chromosome segregation defects		
ZK1251.1	htas-1	Histone H2A variant	High levels on sperm DNA during and after meiotic divisions	None detected (mutant)	None detected	Partial Ste (mutant)/(RNAi resistant)	Spermatogenesis-specific localization, mutant hermaphrodite has decreased fertility	M, H2AX (2 × 10 ⁻²¹), macroH2A(3 × 10 ⁻²⁷) H, H2AX (3 × 10 ⁻²¹), macroH2A (3 × 10 ⁻²⁷)	M, H2AX KO male infertile
C04G2.8‡	spch-1	SNBP-like	High levels on sperm DNA during and after meiotic divisions	Low penetrance male meiotic problems	None detected	Low Ste/(RNAi resistant)	Spermatogenesis-specific localization	M, histone H1T2 (2.3 × 10 ⁻¹⁴) H, SON DNA-binding protein (1.3 × 10 ⁻¹⁹)	M, Histone H1T2 KO male infertile
C10G11.9‡	spch-2	SNBP-like	High levels on sperm DNA during and after meiotic divisions	None detected	None detected	(RNAi resistant)	Spermatogenesis-specific localization	M, histone H1T2 (2.3 × 10 ⁻¹⁴) H, SON DNA-binding protein (1.3 × 10 ⁻¹⁹)	M, Histone H1T2 KO male infertile
T27A3.4‡	spch-3	SNBP-like	High levels on sperm DNA during and after meiotic divisions	None detected	None detected	(RNAi resistant)	Spermatogenesis-specific localization	M, histone H1T2 (2.3 × 10 ⁻¹⁴) H, SON DNA-binding protein (6.5 × 10 ⁻²⁰)	M, Histone H1T2 KO male infertile
Category II									
M01E5.5	top-1	Topoisomerase I ⁹	Sperm: high levels around mature sperm DNA Other: nucleolar, absent during meiosis	Abnormal gonad morphology, meiotic problems with large nuclei in later germ line	Abnormal gonad morphology, oogenesis arrest or Emo	Complete F ₁ sterility, Gro	Localization around mature sperm chromatin	M, TOP-1 (3 × 10 ⁻¹⁵⁴) H, TOP-1 (1 × 10 ⁻¹⁵⁶)	M, TOP-1 KO Embryonic lethal H, decreased Top-1 activity in testes of infertile men with varicocele
C55B7.1	glh-2	RNA helicase ¹⁰	Sperm: surrounding sperm meiotic chromosomes Other: P-granule localization in germ line	None detected	None detected	WT	Localization around sperm meiotic chromatin	M, Ddx4 (VASA) (1 × 10 ⁻¹⁰⁵) H, DBY (1 × 10 ⁻⁹²), DDX4 (1 × 10 ⁻¹⁰⁵)	M, Ddx4 KO male infertile H, DBY is frequently deleted in infertile patients
T03F1.9	hcp-4	CENP-C centromere component ¹¹	Sperm: surrounding sperm chromosomes during and after meiotic divisions Other: uniformly distributed on oocyte meiotic chromosomes, polewards on mitotic chromosomes	No F ₁ progeny	No F ₁ progeny	Complete F ₁ lethality, Emb	Localization around mature sperm chromatin	M, CENP-C ¹¹ H, CENP-C ¹¹	

The categories are described in the text. Degrees of sterility are defined in Supplementary Methods. Symbols (*, †, ‡) denote highly identical genes whose products may be depleted simultaneously by RNAi. Mammalian fertility link references are listed in Supplementary Table 8. RNAi resistant means that the protein product was detectable by immunolocalization in animals subjected to RNAi of the corresponding gene. Extent of homology is indicated in parenthesis (BLAST e-values) in 'Known homologue' column. Emb, embryonic lethality; Emo, endomitotic reduplication; Gro, slow growing; H, human homologues; KO, mouse knockout of corresponding gene; M, mouse homologues; Ste, sterility; WT, wild type.

homologues not yet tested for roles in fertility (Table 1 and Supplementary Table 9). Analysis of these proteins in mice and humans has the potential to define new mammalian fertility factors. These factors present opportunities for the development of diagnostic tests to assess sperm competence and human reproductive potential, and represent potential targets for the development of safe male contraceptives.

METHODS

Purification of spermatogenic and oogenic chromatin. Germ-cell isolation and chromatin purification are described in Supplementary Methods; 5–20% of spermatogenic nuclei were in various stages of meiosis as judged by a cytological examination of 4',6-diamidino-2-phenylindole (DAPI)-stained chromosomes. Similarly, 10–30% of *fer-1(hc1)* oocytes were in diakinesis, 50–70% were oocytes that had undergone endomitotic reduplication (Emo), and 20–30% were fertilized embryos, allowing the subtraction of chromatin and meiotic factors. Subsequent SDS–PAGE and staining with colloidal blue (Novex) of purified chromatin revealed that core histones were the most abundant proteins in both cell types, showing similar enrichment of chromatin proteins (Supplementary Fig. 2). Major sperm proteins (MSPs), the most abundant proteins in sperm cytosol and pseudopods²⁹, were undetectable in our chromatin preparations by staining with colloidal blue, indicating that sperm cellular components had been effectively removed.

Proteomic identification and subtractive analysis. A total of eleven 12-step LC–LC–MS–MS experiments (six of spermatogenic chromatin and five of oogenic chromatin) were performed to enhance the detection of peptides from small proteins and low-abundance proteins. For each experiment, 50–100 µg of protein was digested and analysed (Supplementary Methods). To ensure the subsequent subtraction of low-abundance oogenic factors, more stringent criteria were adopted for identifying proteins from spermatogenic germ cells (a minimum of two different peptides per protein from each individual preparation) than from oogenic cells (a minimum of two different peptides per protein from all data sets of five preparations combined).

Prioritization of enriched and reliably detected proteins by using two abundance measurements from MudPIT reduced the number of proteins to a manageable subset for functional studies. The first, total spectrum count (TSC), represents the total number of tandem mass spectra collected for each protein from all preparations and was used to assess relative protein abundance³⁰. The second measure, 'occurrence', was the number of preparations in which each protein was identified and indicates the reliability of protein detection across different preparations. Occurrence is especially useful for proteins that are not sampled frequently in any one mass spectrometric analysis. By choosing only spermatogenic proteins represented by three or more occurrences, we reduced the number of candidate proteins from 1,099 to 502. These well-represented spermatogenic proteins comprise 46% (502 of 1,099) of individual proteins identified by MudPIT analysis and constitute 88% of the mass of total proteins in spermatogenic preparations, as estimated by TSC (Supplementary Table 5a). Enrichment of proteins such as core histones was confirmed by high TSC and high occurrence (6 of 6 preparations; Supplementary Fig. 2 and Supplementary Table 3). Correspondingly, MSPs were detected in only very low abundance (Supplementary Tables 2 and 3).

The entire oogenic protein list (812) was subtracted from the prioritized spermatogenic protein list (502). Subtractive analysis removed appropriate factors and reduced the number of proteins to 132 (Supplementary Methods).

Functional analysis. RNAi analysis was conducted for genes corresponding to each of the 132 abundant sperm chromatin proteins by injecting 1–4 mg ml⁻¹ double-stranded RNA into *him-8(e1849)* animals. Quantitative screening and statistical analysis of F₂ animals from RNAi-treated grandparents were used to assess the functions of each gene in fertility (Supplementary Methods). F₁ parents producing statistically lower progeny numbers were further analysed by observing the morphology of germ-cell chromosomes (Supplementary Methods). To assess male fertility, F₁ male progeny were tested for their ability to rescue the fertilization defect of *spe-8 dpy-4* hermaphrodites.

Received 21 April; accepted 6 July 2006.

Published online 30 August 2006.

- Agarwal, A. & Said, T. M. Role of sperm chromatin abnormalities and DNA damage in male infertility. *Hum. Reprod. Update* **9**, 331–345 (2003).
- Kimmins, S. & Sassone-Corsi, P. Chromatin remodelling and epigenetic features of germ cells. *Nature* **434**, 583–589 (2005).
- Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
- Eng, J. K., McCormack, A. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

- Tabb, D. L., McDonald, W. H. & Yates, J. R. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26 (2002).
- Schirmer, E. C., Florens, L., Guan, T., Yates, J. R. & Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301**, 1380–1382 (2003).
- Skop, A. R., Liu, H., Yates, J., Meyer, B. J. & Heald, R. Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* **305**, 61–66 (2004).
- Browning, H. & Strome, S. A sperm-supplied factor required for embryogenesis in *C. elegans*. *Development* **122**, 391–404 (1996).
- Lee, M. H., Park, H., Shim, G., Lee, J. & Koo, H. S. Regulation of gene expression, cellular localization, and *in vivo* function of *Caenorhabditis elegans* DNA topoisomerase I. *Genes Cells* **6**, 303–312 (2001).
- Gruidl, M. E. *et al.* Multiple potential germ-line helicases are components of the germ-line-specific P granules of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 13837–13842 (1996).
- Moore, L. L. & Roth, M. B. HCP-4, a CENP-C like protein in *Caenorhabditis elegans*, is required for resolution of sister centromeres. *J. Cell Biol.* **153**, 1199–1208 (2001).
- Reinke, V., Gil, I., Ward, S. & Kazmer, K. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* **131**, 311–323 (2004).
- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Sonnichsen, B. *et al.* Full genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462–469 (2005).
- Varmuza, S. *et al.* Spermiogenesis is impaired in mice bearing a targeted mutation in the protein phosphatase 1c gamma gene. *Dev. Biol.* **205**, 98–110 (1999).
- Boag, P. R., Ren, P., Newton, S. E. & Gasser, R. B. Molecular characterisation of a male specific serine/threonine phosphatase from *Oesophagostomum dentatum* (Nematoda: Strongylida), and functional analysis of homologues in *Caenorhabditis elegans*. *Int. J. Parasitol.* **33**, 313–325 (2003).
- Hanazawa, M., Mochii, M., Ueno, N., Kohara, Y. & Iino, Y. Use of cDNA subtraction and RNA interference screens in combination. *Proc. Natl Acad. Sci. USA* **98**, 8686–8691 (2001).
- Rogers, E., Bishop, J. D., Waddle, J. A., Schumacher, J. M. & Lin, R. The aurora kinase AIR-2 functions in the release of chromosome cohesion in *Caenorhabditis elegans* meiosis. *J. Cell Biol.* **157**, 219–229 (2002).
- Celeste, A. *et al.* Genomic instability in mice lacking histone H2AX. *Science* **296**, 922–927 (2002).
- Lewis, J. D., McParland, R. & Ausio, J. PL I of *Spisula solidissima*, a highly elongated sperm specific histone H1. *Biochemistry* **43**, 7766–7775 (2004).
- Lewis, J. D., Song, Y., de Jong, M. E., Bagha, S. M. & Ausio, J. A walk through vertebrate and invertebrate protamines. *Chromosoma* **111**, 473–482 (2003).
- Martianov, I. *et al.* Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. *Proc. Natl Acad. Sci. USA* **102**, 2808–2813 (2005).
- Swanson, W. J. & Vacquier, V. D. The rapid evolution of reproductive proteins. *Nature Rev. Genet.* **3**, 137–144 (2002).
- Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Rev. Mol. Cell Biol.* **3**, 430–440 (2002).
- Rossi, F. *et al.* Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* **381**, 80–82 (1996).
- Kawano, T., Fujita, M. & Sakamoto, H. Unique and redundant functions of SR proteins, a conserved family of splicing factors, in *Caenorhabditis elegans* development. *Mech. Dev.* **95**, 67–76 (2000).
- Tanaka, S. S. *et al.* The mouse homolog of *Drosophila* Vasa is required for the development of male germ cells. *Genes Dev.* **14**, 841–853 (2000).
- Vogt, P. H. Azoospermia factor (AZF) in Yq11: towards a molecular understanding of its function for human male fertility and spermatogenesis. *Reprod. Biomed. Online* **10**, 81–93 (2005).
- Ward, S. & Klass, M. The location of the major protein in *Caenorhabditis elegans* sperm and spermatocytes. *Dev. Biol.* **92**, 203–208 (1982).
- Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance. *Anal. Chem.* **76**, 4193–4201 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Strome, K. Bennett, H.-S. Koo, L. Moore, E. Shulze and J. Aris for providing antibodies; A. Villeneuve, G. Stanfield, S. Mitani and the *C. elegans* Genetic Center (CGC) for providing strains; V. Reinke, L. Moore and R. Navarro for sharing unpublished data; D. King for peptide synthesis; A. Chan for help with microscopy; S. Chu for statistical analysis; A. Severson, T. Cline, A. Skop, E. Xu and J. Gladden for comments on the manuscript; and members of the Meyer laboratory for input on this project. This work was funded by the National Institutes of Health grants to D.S.C., J.R.Y. and B.J.M. B.J.M. is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.S.C. (chud@sfsu.edu).

LETTERS

AtSNX1 defines an endosome for auxin-carrier trafficking in *Arabidopsis*

Yvon Jaillais¹, Isabelle Fobis-Loisy¹, Christine Miège¹, Claire Rollin¹ & Thierry Gaudé¹

Polarized cellular distribution of the phytohormone auxin and its carriers is essential for normal plant growth and development. Polar auxin transport^{1,2} is maintained by a network of auxin influx (AUX)³ and efflux (PIN)^{1,2,4–6} carriers. Both auxin transport and PIN protein cycling between the plasma membrane and endosomes require the activity of the endosomal GNOM^{1,2,7,8}; however, intracellular routes taken by these carriers remain largely unknown. Here we show that *Arabidopsis thaliana* SORTING NEXIN 1 (AtSNX1) is involved in the auxin pathway and that PIN2, but not PIN1 or AUX1, is transported through AtSNX1-containing endosomes. We demonstrate that the *snx1*-null mutant exhibits multiple auxin-related defects and that loss of function of AtSNX1 severely enhances the phenotype of a weak *gnom* mutant. In root cells, we further show that AtSNX1 localizes to an endosomal compartment distinct from GNOM-containing endosomes, and that PIN2 accumulates in this compartment after treatment with the phosphatidylinositol-3-OH kinase inhibitor wortmannin or after a gravity stimulus. Our data reveal the existence of a novel endosomal compartment involved in PIN2 endocytic sorting and plant development.

We previously demonstrated that the product of *Brassica oleracea* SNX1, the plant orthologue of human SNX1, interacts with different plant receptor kinases⁹. SNX proteins are cytoplasmic and membrane-associated proteins and are conserved among eukaryotes¹⁰. The mammalian SNX1 and SNX2 are implicated in endocytic sorting and have an essential function in embryonic development. In the genome of *A. thaliana*, we found only three SNX genes, whereas the SNX family in mammals contains about thirty members.

To investigate the function of SNX1 in *Arabidopsis thaliana*, we studied two allelic T-DNA insertion lines, which are null mutants (Supplementary Fig. 1a, b). These *snx1* mutants showed pleiotropic defects. We chose to focus on the well-documented root-tip cell biology system for study of SNX1 function. Compared with wild type, *snx1* mutants had shorter primary roots, produced fewer secondary roots and exhibited an altered root gravitropic response (Supplementary Fig. 2). Genomic DNA fragments spanning the *AtSNX1* gene fused to complementary DNAs encoding green (GFP6) or monomeric red (mRFP1)¹¹ fluorescent protein were able to complement *snx1* (Supplementary Figs 1 and 3), indicating that *AtSNX1* is the only gene impaired in the mutant lines. The features of the *snx1* phenotype resembled those of the *gnom*^{B/E} mutant, a very weak allele of the *GNOM* gene⁷. *GNOM* encodes a GDP/GTP exchange factor for small G-proteins (ARF-GEF) and is required for the proper polar localization of PIN1 (refs 1, 7, 8). Notably, like weak *gnom* alleles, auxin maxima were disturbed in the roots of *snx1* mutants (Supplementary Fig. 4). Moreover, *snx1 gnom*^{B/E} double mutant plants exhibited a more severe phenotype than either of the single mutants (Supplementary Fig. 5). The marked

phenotype of double mutant plants is not reminiscent of that caused by stronger *gnom* alleles, for example, *gnom*^{R5} (ref. 7). This indicates that AtSNX1 and GNOM are involved in distinct pathways, but nonetheless have partially overlapping functions in plant growth and development.

To gain insight into the intracellular localization of AtSNX1, we used the complemented *snx1-1* transgenic lines expressing functional AtSNX1–GFP or AtSNX1–mRFP proteins. In both lines, we observed fluorescent AtSNX1 proteins in round-shaped intracellular compartments (Fig. 1b, n). We crossed the *AtSNX1-mRFP* line with marker lines that constitutively express GFP-tagged proteins targeted to different intracellular compartments. AtSNX1–mRFP did not co-localize with two Golgi markers, ERD2–GFP (ref. 12; Fig. 1a–c) and ST–GFP (ref. 12; data not shown), or with TIG2a–GFP (refs 1, 13, 14; Fig. 1d–f), which labels *trans*-Golgi networks in plant cells. Then, we crossed the *AtSNX1-mRFP* line with two lines expressing either the RABF1–GFP or the GFP–RABF2b endosomal markers¹⁵. We observed that AtSNX1–mRFP co-localized with both RABF GFP-tagged proteins in *Arabidopsis* roots (Fig. 1g–l). These results indicate that AtSNX1 is located in an endosomal compartment. To verify this prediction, we used the fluorescent dye FM4-64 as an endocytic tracer^{1,14,16}. In roots of *AtSNX1-GFP* plants, AtSNX1–GFP-containing compartments were labelled after only 15 min of FM4-64 internalization (Fig. 1m–o). Next, we treated roots with brefeldin A (BFA), which inhibits transport from endosomes to the plasma membrane and causes endosomes to aggregate into BFA compartments^{1,2,16}. After 30 min of BFA treatment, AtSNX1–GFP labelling accumulated with endocytosed FM4-64 in BFA compartments (Fig. 1p–r). Altogether, our results indicate that AtSNX1-labelled compartments are endosomes.

As GNOM is an endosomal BFA target in *Arabidopsis* root cells¹, we presumed that AtSNX1 and GNOM are located in the same endosomes. To verify this, we crossed a *GNOM-GFP* line¹ with the *AtSNX1-mRFP* line. No co-localization between the two fluorescent markers was detected (Fig. 2a–c), although both were found in BFA-sensitive compartments (Fig. 2d–f). Inhibitors of phosphatidylinositol-3-OH kinase, such as wortmannin, affect SNX localization in mammalian cells^{17,18}. In root cells, wortmannin treatment led to the appearance of one or two enlarged, vacuolated AtSNX1-containing compartments per cell, but did not alter the morphology of GNOM-containing endosomes (Fig. 2g–i and Supplementary Fig. 6). We conclude that GNOM- and AtSNX1-containing endosomes are physically distinct; although both are BFA sensitive, only AtSNX1-containing endosomes exhibit sensitivity to wortmannin.

Although GNOM regulates trafficking of the basally localized PIN1, recycling of PIN2 is partially independent of GNOM¹. This implies that additional endocytic routes are involved in recycling of PIN2 and we proposed that AtSNX1 endosomes might be one of

¹Reproduction et Développement des Plantes, IFR128 BioSciences Lyon-Gerland, Centre National de la Recherche Scientifique, Institut National de la Recherche Agronomique, Université Claude Bernard Lyon I, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France.

those. To test this assumption, we treated seedlings expressing functional PIN1–GFP (ref. 19) and PIN2–GFP (ref. 20) with wortmannin. The wortmannin treatment was concomitant with an incubation of the same roots in the protein synthesis inhibitor cycloheximide, to ensure that GFP labelling reveals only previously synthesized proteins. Whereas no difference was observed in the localization of PIN1–GFP after wortmannin treatment (Fig. 2j, k; as expected because only the GNOM route is involved in PIN1 recycling), some PIN2–GFP was relocalized to wortmannin-induced compartments (Fig. 2l, m). Moreover, PIN2–GFP that remained at the plasma membrane seemed appropriately polarized, apically in epidermal cells and basally in cortical cells. In addition, the epidermal cells, compared with cortical cells, were particularly sensitive to wortmannin, indicating that PIN2 trafficking might be differentially regulated in these two distinct cell files. The wortmannin-induced compartments were shown to co-localize with AtSNX1–mRFP in hybrid transgenic plants co-expressing PIN2–GFP (Fig. 2n–p). Stimulation by gravity (gravistimulation) of root apices has recently been shown to result in internalization of PIN2 proteins at the upper side of the root, whereas endocytosis is inhibited at the lower side^{14,21}. To explore whether PIN2 internalization during gravistimulation involves AtSNX1 endosomes, we submitted seedlings of the PIN2–GFP/AtSNX1–mRFP hybrid plant to a gravity

stimulus. Endocytosed PIN2–GFP co-localized with AtSNX1–mRFP compartments, indicating that PIN2 transits through these endosomes (Fig. 2q–s). AUX1 is another auxin carrier involved in root gravitropism³ that might be routed via AtSNX1 endosomes. To test

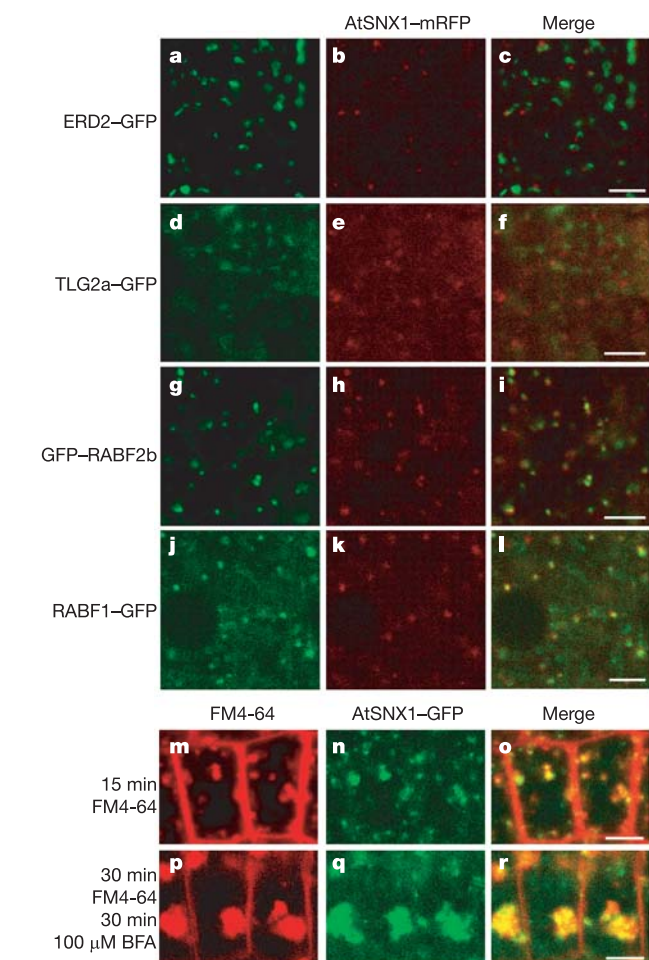


Figure 1 | AtSNX1 is localized to an endosomal compartment. **a–l**, Root tips co-expressing AtSNX1–mRFP and a GFP-tagged protein labelling the Golgi apparatus (ERD2–GFP, **a–c**), *trans*-Golgi network (TLG2a–GFP, **d–f**) or endosomes (GFP–RABF2b, **g–i**, and RABF1–GFP, **j–l**). **m–o**, Root tips of AtSNX1–GFP seedlings after 15 min of FM4–64 uptake. **p–r**, BFA-treated root tips of AtSNX1–GFP seedlings after 30 min of FM4–64 uptake. Scale bars, 5 μ m.

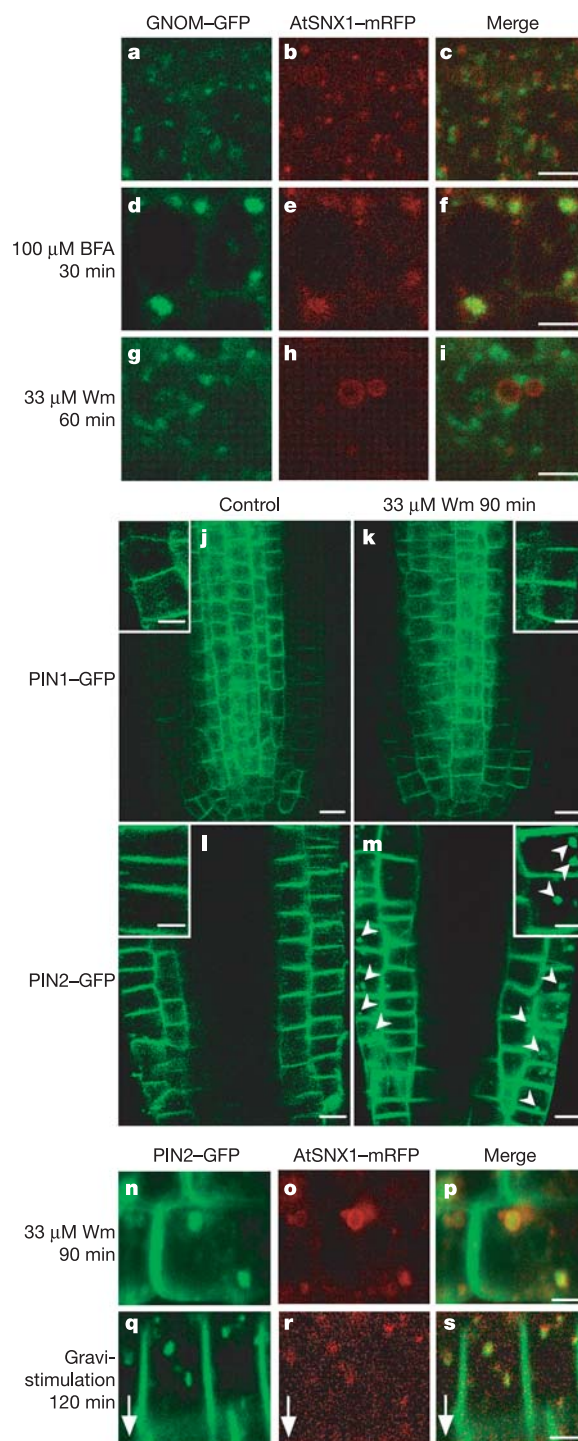


Figure 2 | AtSNX1 defines a new endosomal compartment that transports PIN2. **a–i**, Root tips co-expressing GNOM–GFP and AtSNX1–mRFP without treatment (**a–c**), after 30 min of BFA treatment (**d–f**) and after 60 min of wortmannin (Wm) treatment (**g–i**). **j–m**, PIN1–GFP (**j–k**) and PIN2–GFP (**l–m**) localization in root tips after 90 min of cycloheximide alone or cycloheximide and wortmannin treatment. **n–s**, Root tips co-expressing PIN2–GFP and AtSNX1–mRFP after 90 min of cycloheximide and wortmannin treatment (**n–p**), or after 120 min of gravistimulation (**q–s**). Arrowheads indicate wortmannin-induced compartments; arrows indicate the gravity vector. Scale bars: 5 μ m (**a–i**; **n–s**); 10 μ m (**j–m**).

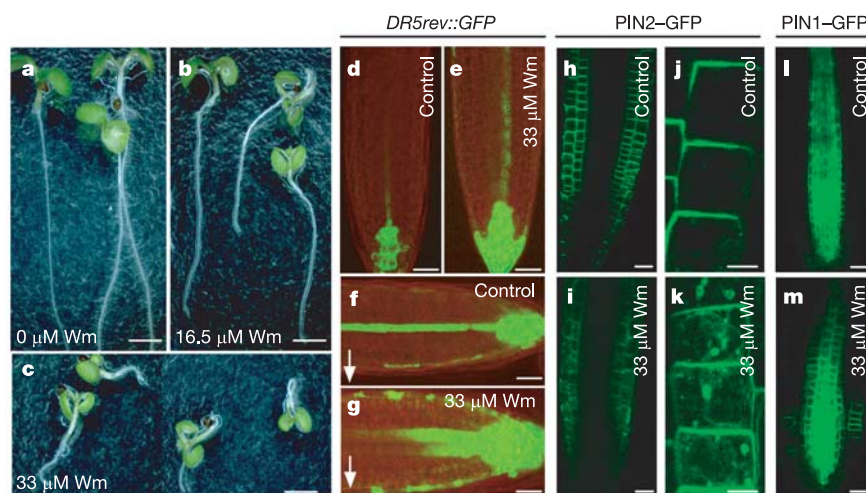


Figure 3 | Wortmannin effects on seedling development, auxin distribution and PIN localization. Five-day-old seedlings vertically grown on MS medium alone (**a, d, f, h, j, l**) or supplemented with 16.5 μ M wortmannin (**b**) or 33 μ M wortmannin (**c, e, g, i, k, m**). **a–c**, Wild-type seedlings. **d–g**, *DR5rev::GFP* expression without gravistimulation (**d, e**) or after

180 min of gravistimulation (**f, g**). Arrows indicate the gravity vector. **h–m**, PIN2-GFP (**h–k**) and PIN1-GFP (**l, m**) localization. Scale bars: 5 mm (**a–c**), 20 μ m (**d–i, l, m**), 5 μ m (**j, k**). The same laser intensity was used in **d, e, f, g, h, i, j, k, l, m**.

this possibility, we treated roots of an *AUX1-YFP* line²² separately with BFA and wortmannin. Under our experimental conditions, neither drug affected *AUX1-YFP* localization (Supplementary Fig. 7). Localization of GFP or YFP fusion auxin carriers has been shown to correlate with that of the native proteins^{23,24}. It is reasonable to consider that the behaviour of fluorescent proteins in our experiments reflects the genuine dynamics of auxin carriers, although we cannot exclude that auxin carriers fused to GFP or YFP might have altered stability. Altogether, our results indicate that PIN2, but not PIN1 or *AUX1*, trafficks through wortmannin-sensitive, AtSNX1-containing endosomes.

Next, we investigated the effects of long-term wortmannin treatment on *Arabidopsis* seedlings, to see whether alteration of the AtSNX1 endosomal compartment (Supplementary Fig. 8) may have significant consequences for plant development, auxin distribution or PIN2 localization. Phenotypic features reminiscent of those provoked by auxin transport inhibitors², such as defects in root and hypocotyl gravitropism and elongation, were induced by wortmannin in a concentration-dependent manner (Fig. 3a–c and Supplementary Fig. 9). These modifications were accompanied by a clear alteration of auxin distribution, exemplified by the signals obtained using the auxin-maxima reporter *DR5rev::GFP* (refs 5 and 19) in the presence or absence of a gravity stimulus (Fig. 3d–g). The increased GFP signal at the root tip, showing extension into lateral root-cap and epidermis cells (Fig. 3e), was very similar to the *DR5rev::GFP* expression observed in plants with the *eir1-4* (*pin2*) knockout allele²¹. This indicated that wortmannin treatment might affect PIN2 abundance and subcellular location. Consistent with this prediction, abundance and localization of PIN2-GFP were markedly perturbed in cortical and epidermal cells of wortmannin-treated seedlings (Fig. 3h–k). Only a faint fluorescent signal was observed in root tips of wortmannin-treated seedlings compared with the control. Although PIN2-GFP polarity was still maintained at the plasma membrane, PIN2-GFP accumulated in wortmannin-induced compartments and also labelled the tonoplast (Fig. 3k). It is probable that endocytosed PIN2-GFP was targeted for degradation, as demonstrated by the weak intensity of the PIN2-GFP signal. In contrast, localization of PIN1-GFP after wortmannin treatment was not markedly different from control seedlings, except that PIN1-GFP expression slightly expanded into some cortical cells (Fig. 3l, m). This is reminiscent of the ectopic expression of *PIN1* in the *PIN2* expression domains reported to occur in *pin2* (ref. 24). The low

levels of PIN2-GFP detected in cortical and epidermal cells under wortmannin treatment could considerably reduce the uptake of auxin from columella cells. This would account for the observed enrichment of auxin in the root tip and the agravitropic phenotype of wortmannin-treated seedlings.

Taken together, our data show that AtSNX1 has a function in root growth and gravitropic response. Although our results are consistent with altered auxin accumulations and indirectly support altered auxin transport in *snx1* root tips, free indole-3-acetic acid levels and auxin transport were not assayed. It is possible that the weak phenotype we observed for *snx1* mutants is the result of a partial redundancy of AtSNX1 with the other two AtSNX proteins (AtSNX2a and AtSNX2b)⁹ or other proteins involved in intracellular trafficking. The discovery of the wortmannin-sensitive, AtSNX1-containing endosomes, and the effect of wortmannin on PIN2-GFP relocalization or degradation in roots, strongly indicate that AtSNX1 defines a novel endosomal compartment involved in auxin-carrier trafficking. On the basis of the endosomal localization of AtSNX1, which differs from that of the GNOM protein, we propose that there are at least two distinct endosomal routes for the recycling or trafficking of auxin transport components: one involving GNOM-containing endosomes (used by PIN1 and partially by PIN2), and a second route using AtSNX1-containing endosomes (used by PIN2, but not PIN1 or *AUX1*). This dual option of regulating polar auxin trafficking may ensure a fine control of auxin levels and auxin carriers at the subcellular level.

In the past few years, several studies in developmental biology using animal systems have brought to light a complex and dynamic picture of the endosomal system²⁵. Our data underline the importance of endosomal proteins in the control of growth and developmental processes in plants. Identification of the other cargoes and polar auxin transport components routed by the AtSNX1-containing endosomes, and determination of developmental pathways that they regulate will be the next challenge for the coming years.

METHODS

Material and growth conditions. The *snx1-1* T-DNA mutant (Columbia accession, GABI-Kat_105C01) was provided by B. Weisshaar and the *snx1-2* T-DNA mutant (Columbia accession, SALK_03351) was obtained from the SALK Institute²⁶. The *pGNOM:GNOM-GFP* line in a *gnom* background, *pPIN1:PIN1-GFP* line in a *pin1* background, *pPIN2:PIN2-GFP* line in a *pin2* (*eir1-1*) background, *pAUX1:AUX1-YFP* line, *DR5::GUS* line, *DR5rev::GFP* line,

and the *gnom*^{B/E} mutant were described previously^{1,5,7,19,20,22,27}. Plants were grown on soil with long-day conditions at 21 °C, 70% humidity. For gravistimulation experiments, seedlings were grown vertically for 5–7 days in Petri dishes and analysed as described⁷.

Microscopy and drug treatments. Roots of 7-day-old seedlings grown on Murashige and Skoog (MS) medium were mounted in LM medium¹⁶ and analysed on a LSM-510 laser-scanning confocal microscope (Zeiss). For FM4-64 treatments, seedlings were incubated for 2 min at room temperature in LM containing FM4-64 (3.5 µM) and rinsed three times in LM before observation. BFA (100 µM in DMSO/ethanol), wortmannin (33 µM in DMSO) and cycloheximide (50 µM in DMSO) treatments were also performed in LM, for the times indicated, before observation. Recombinant plasmid construction, plant transformation and statistical analysis of results are described in Supplementary Information.

Received 11 April; accepted 5 July 2006.

Published online 27 August 2006.

- Geldner, N. *et al.* The *Arabidopsis* GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. *Cell* **112**, 219–230 (2003).
- Geldner, N., Friml, J., Stierhof, Y. D., Jürgens, G. & Palme, K. Auxin transport inhibitors block PIN1 cycling and vesicle trafficking. *Nature* **413**, 425–428 (2001).
- Swarup, R. *et al.* Localization of the auxin permease AUX1 suggests two functionally distinct hormone transport pathways operate in the *Arabidopsis* root apex. *Genes Dev.* **15**, 2648–2653 (2001).
- Friml, J., Wiśniewska, J., Benková, E., Mendgen, K. & Palme, K. Lateral relocation of auxin efflux regulator PIN3 mediates tropism in *Arabidopsis*. *Nature* **415**, 806–809 (2002).
- Friml, J. *et al.* Efflux-dependent auxin gradients establish the apical–basal axis of *Arabidopsis*. *Nature* **426**, 147–153 (2003).
- Petrásek, J. *et al.* PIN proteins perform a rate-limiting function in cellular auxin efflux. *Science* **312**, 914–918 (2006).
- Geldner, N. *et al.* Partial loss-of-function alleles reveal a role for GNOM in auxin transport-related, post-embryonic development of *Arabidopsis*. *Development* **131**, 389–400 (2004).
- Steinmann, T. *et al.* Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science* **286**, 316–318 (1999).
- Vanoosthuysen, V., Tichtinsky, G., Dumas, C., Gaude, T. & Cock, J. M. Interaction of calmodulin, a sorting nexin and kinase-associated protein phosphatase with the *Brassica oleracea* S locus receptor kinase. *Plant Physiol.* **133**, 919–929 (2003).
- Carlton, J., Bujny, M., Rutherford, A. & Cullen, P. Sorting nexins—unifying trends and new perspectives. *Traffic* **6**, 75–82 (2005).
- Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).
- Brandizzi, F., Snapp, E. L., Roberts, A. G., Lippincott-Schwartz, J. & Hawes, C. Membrane protein transport between the endoplasmic reticulum and the Golgi in tobacco leaves is energy dependent but cytoskeleton independent: evidence from selective photobleaching. *Plant Cell* **14**, 1293–1309 (2002).
- Bassham, D. C., Sanderfoot, A. A., Kovaleva, V., Zheng, H. & Raikhel, N. V. AtVPS45 complex formation at the *trans*-Golgi network. *Mol. Biol. Cell* **11**, 2251–2265 (2000).
- Paciorek, T. *et al.* Auxin inhibits endocytosis and promotes its own efflux from cells. *Nature* **435**, 1251–1256 (2005).
- Ueda, T., Uemura, T., Sato, M. H. & Nakano, A. Functional differentiation of endosomes in *Arabidopsis* cells. *Plant J.* **40**, 783–789 (2004).
- Grebe, M. *et al.* *Arabidopsis* sterol endocytosis involves actin-mediated trafficking via ARA6-positive early endosomes. *Curr. Biol.* **13**, 1378–1387 (2003).
- Xu, Y., Hortsman, H., Seet, L., Wong, S. H. & Hong, W. SNX3 regulates endosomal function through its PX-domain-mediated interaction with PtdIns(3)P. *Nature Cell Biol.* **3**, 658–666 (2001).
- Zhong, Q. *et al.* Endosomal localization and function of sorting nexin 1. *Proc. Natl Acad. Sci. USA* **99**, 6767–6772 (2002).
- Benkova, E. *et al.* Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115**, 591–602 (2003).
- Xu, J. & Scheres, B. Dissection of *Arabidopsis* ADP-RIBOSYLATION FACTOR 1 function in epidermal cell polarity. *Plant Cell* **17**, 525–536 (2005).
- Abas, L. *et al.* Intracellular trafficking and proteolysis of the *Arabidopsis* auxin-efflux facilitator PIN2 are involved in root gravitropism. *Nature Cell Biol.* **8**, 249–256 (2006).
- Swarup, R. *et al.* Structure–function analysis of the presumptive *Arabidopsis* auxin permease AUX1. *Plant Cell* **16**, 3069–3083 (2004).
- Xu, J. *et al.* A molecular framework for plant regeneration. *Science* **311**, 385–388 (2006).
- Vieten, A. *et al.* Functional redundancy of PIN proteins is accompanied by auxin-dependent cross-regulation of PIN expression. *Development* **132**, 4521–4531 (2005).
- Perret, E., Lakkaraju, A., Deborde, S., Schreiner, R. & Rodriguez-Boulán, E. Evolving endosomes: how many varieties and why? *Curr. Opin. Cell Biol.* **17**, 423–434 (2005).
- Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
- Ulmasov, T., Murfett, J., Hagen, G. & Guilfoyle, T. J. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* **9**, 1963–1971 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Max Planck and SALK Institute for providing the insertion mutant lines. We thank C. Dumas for hosting T.G.'s team in the Reproduction et Développement des Plantes laboratory. We acknowledge A. Chaboud for providing *pGlnK::mRFP* plasmids; C. Hawes for providing *ERD2-GFP* and *ST-GFP* plasmids; T. Guilfoyle for providing the *DR5::GUS* transgenic line; G. Jurgens for providing *gnom*^{B/E} mutant and *GNOM-GFP* lines; J. Friml for supplying the *PIN1-GFP* line; B. Scheres for the gift of the *PIN2-GFP* line; M. Bennett for the gift of the *AUX1-YFP* line and NASC for providing the *DR5rev::GFP* line. We are grateful to R. Tsien for the gift of *mRFP1*. We thank C. Scutt for critical reading of the manuscript, N. Paris for helpful discussions and C. Lionnet and F. Simian for technical assistance in confocal microscopy. This work was supported by the Action concertée incitative Biologie du Développement of the French Ministry of Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.G. (thierry.gaude@ens-lyon.fr).

LETTERS

Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex

Carlos Frazão^{1*}, Colin E. McVey^{1*}, Mónica Amblar^{2*}, Ana Barbas², Clemens Vornrhein³, Cecília M. Arraiano² & Maria A. Carrondo¹

RNA degradation is a determining factor in the control of gene expression. The maturation, turnover and quality control of RNA is performed by many different classes of ribonucleases¹. Ribonuclease II (RNase II) is a major exoribonuclease that intervenes in all of these fundamental processes¹; it can act independently or as a component of the exosome, an essential RNA-degrading multiprotein complex². RNase II-like enzymes are found in all three kingdoms of life, but there are no structural data for any of the proteins of this family^{1–5}. Here we report the X-ray crystallographic structures of both the ligand-free (at 2.44 Å resolution) and RNA-bound (at 2.74 Å resolution) forms of *Escherichia coli* RNase II. In contrast to sequence predictions, the structures show that RNase II is organized into four domains: two cold-shock domains, one RNB catalytic domain, which has an unprecedented $\alpha\beta$ -fold, and one S1 domain. The enzyme establishes contacts with RNA in two distinct regions, the ‘anchor’ and the ‘catalytic’ regions, which act synergistically to provide catalysis⁶. The active site is buried within the RNB catalytic domain, in a pocket formed by four conserved sequence motifs. The structure shows that the catalytic pocket is only accessible to single-stranded RNA, and explains the specificity for RNA versus DNA cleavage. It also explains the dynamic mechanism of RNA degradation by providing the structural basis for RNA translocation and enzyme processivity. We propose a reaction mechanism for exonucleolytic RNA degradation involving key conserved residues. Our three-dimensional model corroborates all existing biochemical data for RNase II, and elucidates the general basis for RNA degradation. Moreover, it reveals important structural features that can be extrapolated to other members of this family.

RNase II is a ubiquitous exoribonuclease that processively hydrolyses RNA in the 3′ to 5′ direction, releasing 5′ monophosphates¹. The enzyme binds RNA and in the presence of divalent cations cleaves single-stranded RNA (ssRNA), one nucleotide at a time, independently of sequence³. RNase II is often essential for growth², can be developmentally regulated³, and mutations in its gene have been linked with abnormal chloroplast biogenesis⁴, mitotic control and cancer⁷. In eukaryotes, RNase II (also called Dis3/Rrp44) is a component of both the nuclear and cytoplasmic exosome, a complex of exoRNases crucial for RNA metabolism². The exosome is involved in the maturation and turnover of RNA², RNA interference⁸, and surveillance pathways that recognize and degrade aberrant RNAs^{9–11}. The end-product of exosome degradation, like that of RNase II itself, is a four-nucleotide RNA oligomer^{2,11}. RNase II is also important in the regulation of polyadenylation mediated decay in bacteria¹², and some family members have been shown to be involved in stress responses¹³ and virulence¹⁴.

Sequence homology analysis predicts a similar domain organization for all RNase II-like enzymes that includes a well-conserved catalytic domain, known as RNB (ref. 1), and one or more oligonucleotide-binding domains. The RNB is exclusive to RNase II proteins, and contains four highly conserved motifs (I–IV) with several invariant amino acids¹.

Here we report the crystallographic structures of *E. coli* wild-type RNase II at 2.44 Å resolution, and of a single-amino-acid mutant (D209N) in complex with RNA at 2.74 Å resolution ($R_{\text{work}}/R_{\text{free}}$ is 0.187/0.237 and 0.185/0.224, respectively) (Supplementary Table S1). This is, to our knowledge, the first determined structure of a member of the RNase II family. The structure reveals that RNase II is composed of one RNB domain and three RNA-binding domains, of which one had not been predicted by sequence analysis (Fig. 1a and Supplementary Fig. S1a). The amino-terminal region starts with an α -helix followed by two consecutive five-stranded anti-parallel β -barrels, identified as cold-shock domains (CSD1 and CSD2; Supplementary Table S2 and Supplementary Fig. S8), known to bind RNA^{15,16}. The second CSD (CSD2), which could only be identified by three-dimensional analysis, lacks the typical sequence motifs RNPI and RNPII (ref. 16) but contributes to RNA binding (Fig. 2a, c). At the carboxy terminus, there is a third RNA-binding domain with a typical S1 RNA-binding fold^{5,17} (Supplementary Table S3 and Supplementary Fig. S9). Between the two CSDs and the S1 domain, the RNB catalytic domain shows an unprecedented $\alpha\beta$ -fold (Supplementary Fig. S7). The structure of the RNase II mutant–RNA complex contains a single-stranded 13-nucleotide RNA fragment (modelled as poly(A)) that fortuitously co-purified with the enzyme (Fig. 1a and Supplementary Fig. S1a). This mutant contains a point mutation (D209N) within RNB motif I that allows RNA binding but prevents cleavage¹⁸. A comparison of the free and complexed forms of RNase II shows a conserved overall scaffold with the largest shifts associated with RNA-binding domains (Fig. 1b and Supplementary Fig. S1b).

The RNA fragment interacts with the protein at two non-contiguous regions, usually referred to as ‘anchor’ and ‘catalytic’ regions⁶, which are bridged by a flexible, intermediate region of the RNA chain (Figs 1a, c and 2, and Supplementary Fig. S1a). Nucleotides 1–5, at the 5′-end of the RNA fragment, are located in the anchor region in a deep cleft between the two CSDs and the S1 domain (Fig. 1c, e). They form interactions with the CSD2 and S1 domains, having their phosphates mainly surrounded by solvent and pointing to the top of the cleft, although nucleotides 3–5 define a specific locus of docking (Figs 1e and 2c). CSD1 has no close interactions with the present RNA substrate. Interestingly, most canonical oligonucleotide-binding (OB) motifs¹⁶ in CSD1 and CSD2 face the solvent,

¹Division of Biological Chemistry, ²Division of Biology, ITQB-Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apt. 127, 2781-901 Oeiras, Portugal.

³Global Phasing Limited, Sheraton House, Castle Park, Cambridge CB3 0AX, UK.

*These authors contributed equally to this work.

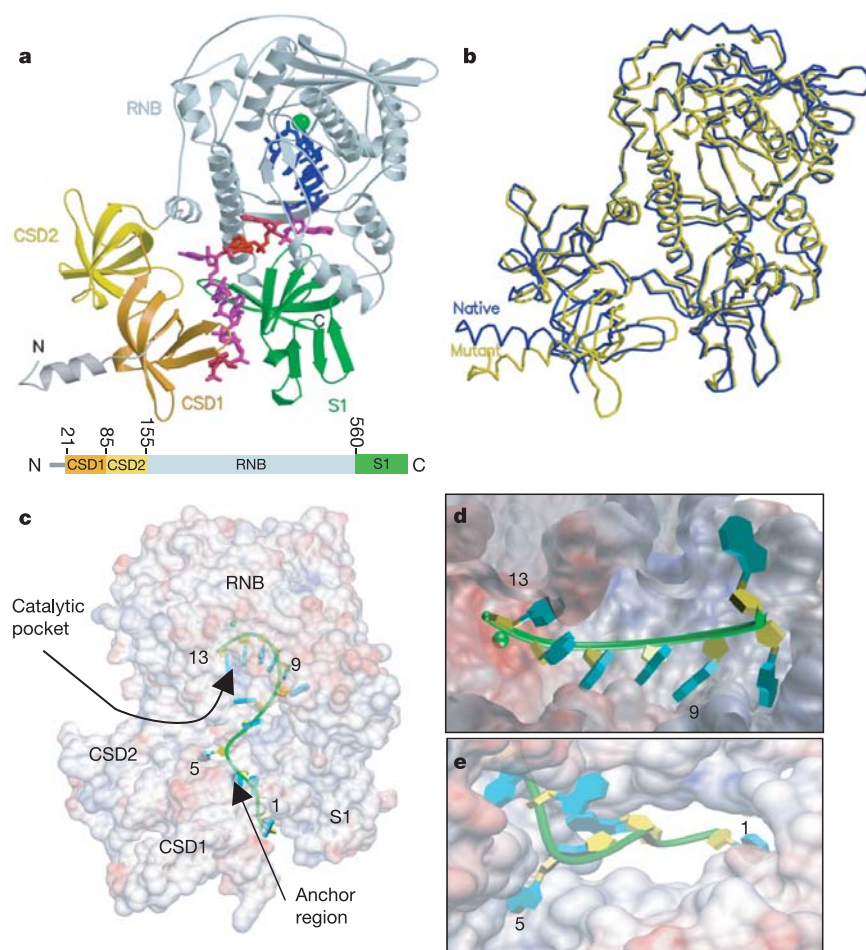


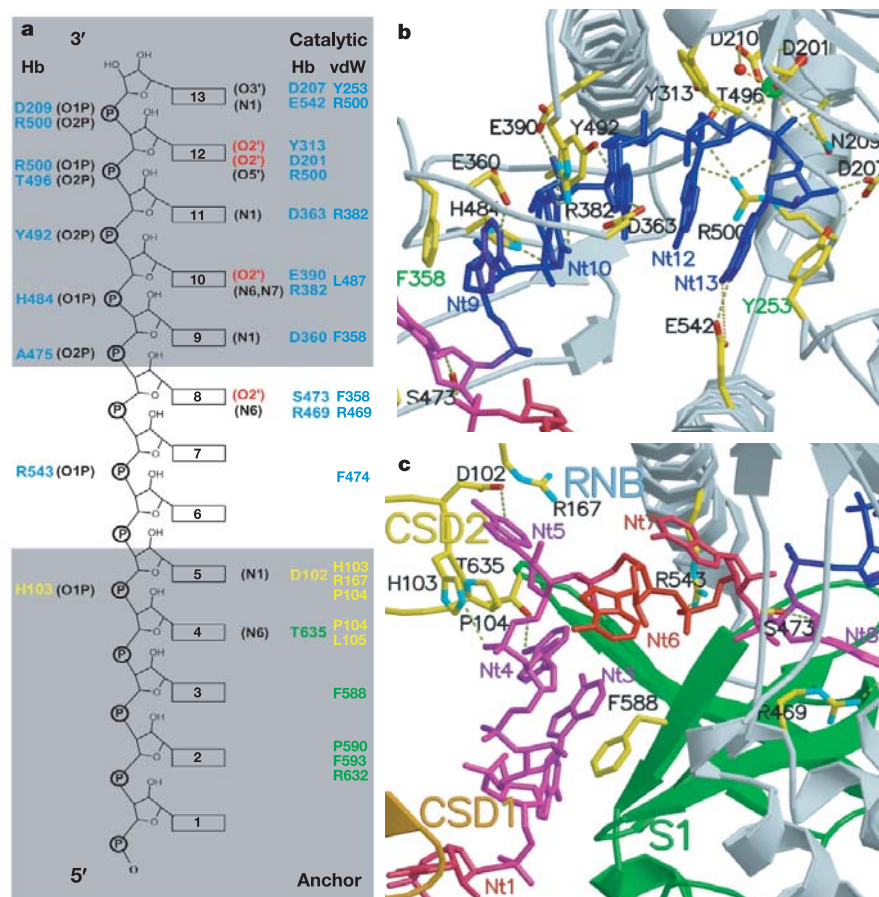
Figure 1 | RNase II and its RNA complex. **a**, Top, representation of RNase II–RNA complex, showing domains CSD1 (orange), CSD2 (yellow), RNB (light cyan) and S1 (green), the Mg ion as a sphere (green) and RNA in sticks (blue to red with increasing B factors). Bottom, linear representation of RNase II domains with numbers specifying the positions of the residues. **b**, Superposition of the polypeptide chains of native RNase II (blue) and its D209N derivative (yellow). **c**, Electrostatic potential of RNase II D209N, mapped on its semi-transparent solvent accessible surface, with docked RNA (phosphate backbone in green, ribose rings in yellow and bases in cyan) with numbers highlighting relevant bases. **d**, **e**, Docking of RNA over the enzyme surface through **d**, the catalytic site (nucleotides 9–13), and **e**, the anchor region (nucleotides 1–5), at the entrance of the cleft. Red to blue surface indicates negative to positive electrostatic potentials.

which may suggest their function in RNA ‘recruitment’. It is possible that this ‘recruitment’ site could substitute for the anchor site to some extent in the context of longer and/or more-structured RNAs (Supplementary Fig. S4). Bridging the anchor to the catalytic region, nucleotides 6–8 are flexible and form a bend in the RNA chain with almost no protein contacts (Figs 1c and 2a). The final nucleotides 9–13 are located in a cavity within the RNB domain (Fig. 1c, d), with their five bases clamped between conserved Phe 358 and Tyr 253 (Fig. 2b and Supplementary Fig. S10). Each phosphate group is engaged in one or two hydrogen bonds with protein residues, a characteristic of non-sequence-specific nucleotide recognition sites. This catalytic cavity is flanked by the four RNase II sequence motifs (Supplementary Fig. S3b). Motifs I and II provide the proper environment for the 3′-end of the substrate where catalysis occurs, motif III contributes to RNase II specificity and motif IV provides interactions with the phosphate backbone. Access to the catalytic pocket is restricted to ssRNA owing to the steric hindrance at its entrance. This feature explains the inability of RNase II to degrade double-stranded RNA (dsRNA). However, this is not a general feature of RNase II enzymes as some members of this family are able to degrade dsRNA, either alone (like RNase R), or with the help of a helicase (Dss1), or as a component of multiprotein complexes².

Although RNase II can bind both DNA and RNA, it can only cleave the latter. It has been demonstrated biochemically that ribose specificity is not for the scissile bond but rather for the nearby nucleotides, and that the presence of ribose between positions 2–5 upstream of the 3′-end is required for RNase II cleavage⁶. The structure reveals that nucleotides 8, 10 and 12 form hydrogen bonds between the O2′-ribose oxygens and side-chains of conserved

residues Glu 390, Asp 201 and Tyr 313 (Fig. 2a, b, and Supplementary Table S4). These interactions seem to be responsible for the proper orientation of RNA at the catalytic cavity, thus allowing cleavage by RNase II. Comparable specificity towards RNA has been identified in other RNases^{19,20}. Previous studies have shown that a minimum of 10–15 nucleotides are required for enzyme processivity⁶. Our data reveal that a 10-nucleotide fragment is the shortest RNA able to retain contacts with both anchor (particularly nucleotides 3–5) and catalytic regions (Fig. 2a). Shorter RNA-fragments establish fewer interactions with the protein, which may explain why the enzyme becomes more distributive. This hypothesis is supported by kinetic studies, which revealed that the presence of two separate binding sites contribute to the processivity of RNase II (ref. 6). A similar mechanism has been proposed for PNPase, a prototype of another major family of exoRNases, in which two separated RNA-binding sites have been found to be responsible for the processive degradation of RNA²¹.

The active site consists of aspartic residues 201, 207, 209 and 210, and Arg 500 (Fig. 3a), all of them conserved in the RNase II family¹ (Supplementary Fig. S3a). One Mg²⁺ ion is found in both structures ligated by Asp 201, Asp 210 and water molecules. In the complex structure, the ion is further coordinated by O3′ of nucleotide 12, and O1P of the leaving nucleotide 13, in a distorted octahedral environment (Fig. 3b and Supplementary Table S5). In the wild-type structure, the side-chain of conserved Arg 500 is disordered (there is no significant electron density after C β), whereas in the complex it is stabilized through interactions with the phosphate backbone of nucleotides 12 and 13 (Fig. 3a). Arg 500 may therefore assist in catalysis by fixing the phosphodiester bond at the cleavage position. In addition, Arg 500 together with the Mg²⁺ ion, owing to their

**Figure 2 | RNase II-ssRNA interactions.**

a, Atomic interactions scheme between RNA (enclosed atoms in black or in red for O2'-protein interactions) and protein residues (coloured by domains as in Fig. 1a). Hb indicates hydrogen bonds up to 3.2 Å, and vdW indicates van der Waals interactions up to 3.6 Å. **b**, **c**, RNA docking catalytic region (**b**) with residues 364–381 deleted for clarity, and anchor region (**c**) involving canonical oligonucleotide-binding motifs¹⁶ L₂₃ of CSD2, and chain β3 and loop L₄₅ of S1. RNA (coloured as in Fig. 1) and protein hydrogen-bonding side-chain residues shown in sticks (carbon in yellow, nitrogen in cyan, oxygen in red), enzyme domains as cartoons (coloured as in Fig. 1). Hydrogen bonds shown as dashed green lines, Mg shown as green sphere with coordination in dashed yellow lines. Nucleotides (Nt) 9–13 are clamped between the aromatic side-chains of Tyr 253 and Phe 358 (with labels highlighted in green), nucleotides 3 and 4 are between Phe 588 and Pro 104, and nucleotide 5 is nestled between His 103, Pro 104, Asp 102 and Arg 167.

positive charges, enhance the susceptibility of the nucleotide-13 phosphorus atom to a nucleophilic attack by a hydrolytic water. A proton donor is required by the ribose hydroxyl group upon splitting of the O3'-P bond, in order to stabilize the transition state of the reaction²². Such protonation could be obtained from the neighbouring ribose 2'-OH or from Wat1 coordinated to Mg²⁺ (Fig. 3a), as both are accessible to protons, either directly from solvent waters or via the OH of Tyr 313 (Figs 2b and 3d). Moreover, Asp 209 must play an essential role in the reaction because its substitution by Asn leads to the total inactivation of the enzyme¹⁸. However, the leaving phosphate group hinders any interaction between the 209 side-chain and the leaving O3' group (Fig. 3a), thus excluding its function as a proton donor in the S_N2-like (bimolecular nucleophilic substitution) mechanism, and strongly suggesting its assistance in another catalytic role (see below).

The enzymatic catalysis of nucleic acid cleavage usually depends on divalent cations²³. In particular, a two-metal-ion mechanism was established for RNase H (ref. 19) and suggested for RNase III on the basis of biochemical and structural data²⁰. A comparison between RNase II and RNase H active sites shows that the Mg²⁺ ion and catalytic residues 201, 209 and 210 of RNase II superimpose with the Mg²⁺ ion and a set of three carboxylic groups in RNase H (Fig. 3c). Furthermore, in RNase II there is room for a second metal ion (Fig. 3a), which could assume the function of the second Mg²⁺ in the two-metal-ion nucleotide cleavage mechanism^{19,20,23}. Upon RNA binding, the second Mg²⁺ would be recruited to the right position at the active site, to assist catalysis, as in the case of RNase III (ref. 20). Although our structures only show one Mg²⁺, we postulate a second labile Mg²⁺ ion coordinated by Asp 209 and Asp 207. Such an ion, symmetrically positioned relative to the leaving phosphorus atom²³, would promote a nucleophilic attack by an activated water molecule

(Fig. 3d). The D209N mutation precludes the coordination of the second Mg²⁺, thus preventing the recruitment of the attacking water in the correct orientation, and therefore explaining the mutant inactivity.

After cleavage of the scissile phosphodiester bond, nucleotide 13 must leave the catalytic site for the enzymatic process to continue. The complex structure shows that nucleotide 13 has higher B factors than the previous nucleotide (Supplementary Fig. S11), and that two of its nitrogen atoms are in close proximity to the carboxylic group of Glu 542 (Fig. 3a). This residue could pull nucleotide 13 out of the base-stacked position upon phosphor-ester cleavage, thus facilitating nucleotide elimination.

The structures we report here have unravelled the mechanisms underlying ssRNA-specific binding and cleavage, and shed light on the mechanisms of RNA translocation and enzyme processivity. On the basis of our structural results, we propose a model for RNase II-mediated RNA degradation (Fig. 3e). Single-stranded RNA bends at the interface between the OB and the RNB domains, and is threaded into the catalytic cavity. In this cavity, a tight packing of the five 3'-terminal nucleotides occurs together with numerous RNA-protein hydrogen bonds and lower RNA B factors. After cleavage, the protein promotes the nucleotide elimination by reinforcing an existing hydrogen bond and formation of a new hydrogen bond with Glu 542. As the nucleotide leaves, the RNA translocates and a new phosphodiester bond gets positioned for the next cleavage event. When the RNA is shorter than five nucleotides, the required packing of the bases cannot occur, thus preventing the translocation of the RNA and generating a four-nucleotide fragment as the final product (Supplementary Fig. S4b). This model shows the dynamics of the RNase II-mediated RNA degradation, and opens a new chapter in the comprehension of RNA metabolism.

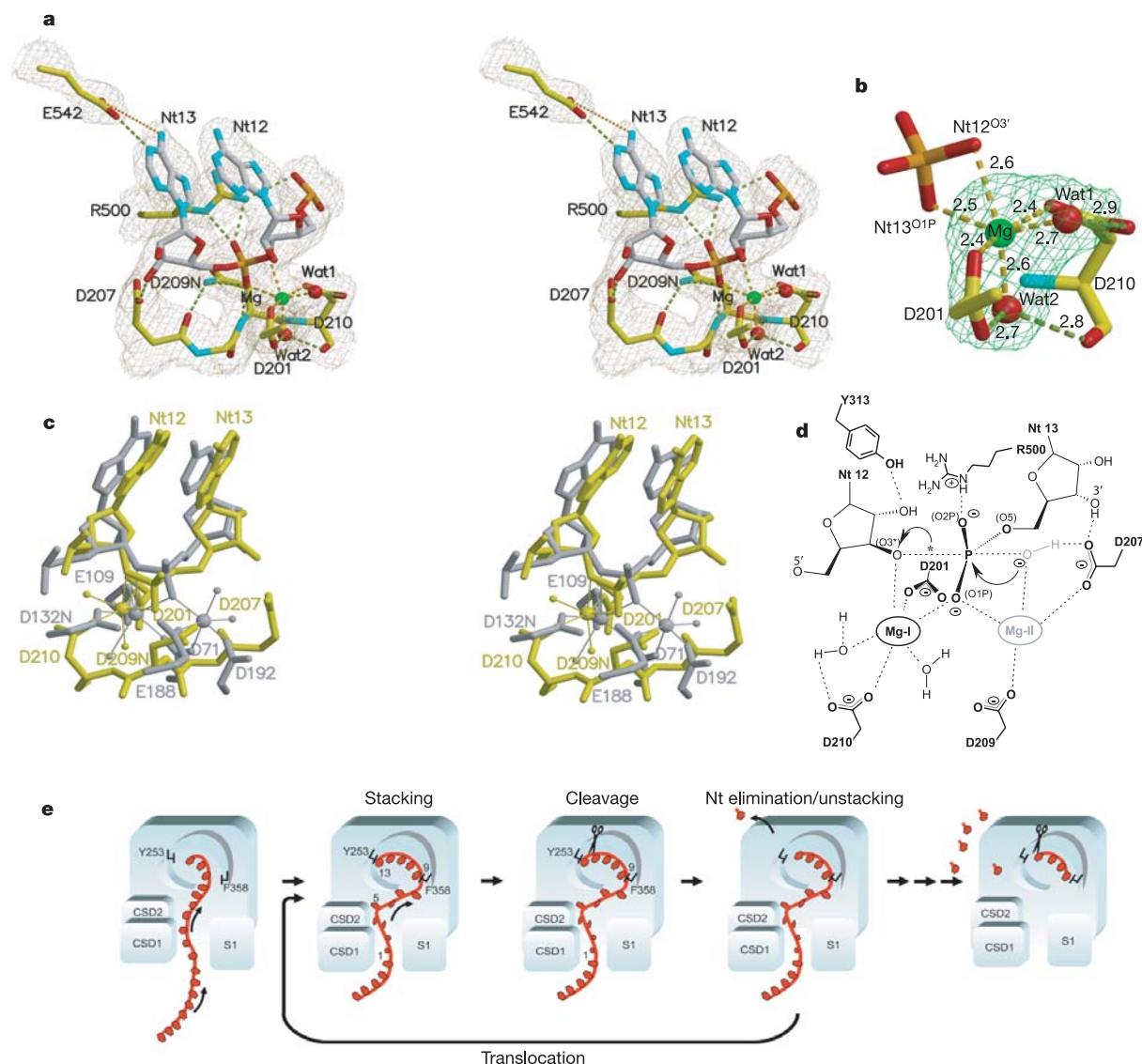


Figure 3 | RNA degradation by RNase II. **a**, Stereo view of RNase II D209N mutant active site; bonds are shown as sticks (oxygen in red, nitrogen in blue, phosphorus in orange), waters as red spheres, Mg as a green sphere and its coordination as dashed orange lines, and hydrogen bonds as dashed green lines, superposed with sigma-A corrected Fourier synthesis electron density map (brown mesh) contoured at 1σ . Additionally, N1 and N6 of nucleotide 13 are in the vicinity of the carboxylate oxygens of Glu 542, at 3.2 Å (hydrogen bond as dashed green lines) and at 4.3 Å (distance as dotted orange line), respectively. **b**, Magnified view of the Mg ion and its coordinating environment (distances in Å) superposed with the positive 3σ sigma-A corrected Fourier difference map (green mesh) calculated with the

Mg and coordinating waters omitted from the model. **c**, Stereo view of the superposition of the RNase II D209N mutant (yellow) and RNase H (grey) with magnesium coordinating spheres (opposite view to Fig. 3a). RNase H displays two magnesium ions ligated by Glu 109, D132N, Glu 188 and Asp 192 that correspond in RNase II to Asp 201, Asp 210, D209N and Asp 207, respectively. **d**, Proposed catalytic mechanism for RNase II, showing the postulated second Mg (Mg II) and the attacking hydroxyl group (grey). **e**, Model for RNA degradation by RNase II. ssRNA (red) is threaded into the catalytic cavity and clamped between Tyr 253 and Phe 358. The additional stabilization of RNA inside the cavity drives the RNA translocation after each cleavage, up to a final four-nucleotide fragment.

METHODS

Details of protein purification, crystallization, diffraction data collection and processing are described elsewhere²⁴. The structure of the mutant complex was solved by MIRAS using selenomethionine and mercury acetate derivatives. Upon solution of the selenium substructure²⁴, SHARP²⁵ was used to locate the Hg sites and to combine and improve the final set of experimental phases. RESOLVE²⁶ was used to obtain initial protein fragments. Model building was performed with XtalView²⁷ followed by refinement with REFMAC²⁸ and iterative rebuilding and model improvement with COOT²⁹. The native protein model was obtained by molecular replacement with MOLREP³⁰ using the mutant protein model as target structure, followed by refinement with REFMAC²⁸.

Received 16 February; accepted 14 July 2006.

1. Mian, I. S. Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res.* **25**, 3187–3195 (1997).

2. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervay, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3' → 5' exoribonucleases. *Cell* **91**, 457–466 (1997).
3. Cairrão, F., Arraiano, C. & Newbury, S. *Drosophila* gene *tazman*, an orthologue of the yeast exosome component Rrp44p/Dis3, is differentially expressed during development. *Dev. Dyn.* **232**, 733–737 (2005).
4. Bollenbach, T. J. et al. RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. *Nucleic Acids Res.* **33**, 2751–2763 (2005).
5. Amblar, M., Barbas, A., Fialho, A. M. & Arraiano, C. M. Characterization of the functional domains of *Escherichia coli* RNase II. *J. Mol. Biol.* **360**, 921–933 (2006).
6. Cannistraro, V. J. & Kennell, D. The processive reaction mechanism of ribonuclease II. *J. Mol. Biol.* **243**, 930–943 (1994).
7. Lim, J. et al. Isolation of murine and human homologues of the fission-yeast *dis3+* gene encoding a mitotic-control protein and its overexpression in cancer cells with progressive phenotype. *Cancer Res.* **57**, 921–925 (1997).

8. Orban, T. I. & Izaurralde, E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**, 459–469 (2005).
9. van Hoof, A., Frischmeyer, P. A., Dietz, H. C. & Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262–2264 (2002).
10. Lejeune, F., Li, X. & Maquat, L. E. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylation, and exonucleolytic activities. *Mol. Cell* **12**, 675–687 (2003).
11. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
12. Dreyfus, M. & Regnier, P. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* **111**, 611–613 (2002).
13. Cairão, F., Cruz, A., Mori, H. & Arraiano, C. M. Cold shock induction of RNase R and its role in the maturation of the quality control mediator SsrA/tmRNA. *Mol. Microbiol.* **50**, 1349–1360 (2003).
14. Cheng, Z. F., Zuo, Y., Li, Z., Rudd, K. E. & Deutscher, M. P. The *vacB* gene required for virulence in *Shigella flexneri* and *Escherichia coli* encodes the exoribonuclease RNase R. *J. Biol. Chem.* **273**, 14077–14080 (1998).
15. Graumann, P. L. & Marahiel, M. A. A superfamily of proteins that contain the cold-shock domain. *Trends Biochem. Sci.* **23**, 286–290 (1998).
16. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133 (2003).
17. Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M. & Murzin, A. G. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* **88**, 235–242 (1997).
18. Amblar, M. & Arraiano, C. M. A single mutation in *Escherichia coli* ribonuclease II inactivates the enzyme without affecting RNA binding. *FEBS J.* **272**, 363–374 (2005).
19. Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016 (2005).
20. Gan, J. *et al.* Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* **124**, 355–366 (2006).
21. Symmons, M. F., Williams, M. G., Luisi, B. F., Jones, G. H. & Carpousis, A. J. Running rings around RNA: a superfamily of phosphate-dependent RNases. *Trends Biochem. Sci.* **27**, 11–18 (2002).
22. Cassano, A. G., Anderson, V. E. & Harris, M. E. Understanding the transition states of phosphodiester bond cleavage: insights from heavy atom isotope effects. *Biopolymers* **73**, 110–129 (2004).
23. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
24. McVey, C. E. *et al.* Expression, purification, crystallization and preliminary diffraction data characterization of *Escherichia coli* ribonuclease II (RNase II). *Acta Crystallogr. F* **62**, 684–687 (2006).
25. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
26. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
27. McRee, D. E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
30. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Maquat, C. Condon and P. Lindley for critical reading of the manuscript; R. Parker, W. Hendrickson and T. Blundell for advice and encouragement; and the staff of beamlines D14-1/ID14-2/ID13 from ESRF, Grenoble, for data collection technical support. This work was supported by Fundação para a Ciência e Tecnologia, Portugal.

Author Contributions C.F., C.E.M. (crystallography) and M.A. (molecular biology) contributed equally to this work.

Author Information Diffraction data and atomic coordinates of RNase II and its D209N-RNA bound mutant complex have been deposited in the Protein Data Bank with accession numbers r2ix0sf and 2ix0, and r2ix1sf and 2ix1, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.C (carrondo@itqb.unl.pt) or C.M.A (cecilia@itqb.unl.pt).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

A report on UK initiatives to transfer skills from academia to industry describes the process as “fragmented, disjointed and not sufficiently robust”. The report to the Science Council, called *The Organization of Science Education, Skills and Knowledge in the UK*, describes 30 initiatives over the past 15 years that used various skills councils to identify industry’s needs and then find ways of encouraging universities to meet them. With a few exceptions, those efforts were fairly dismal, with several surveyed participants unable to cite examples that made a significant impact on performance or profitability.

The report indirectly points to the responsibilities of universities, industry and young scientists. Although research-oriented universities aren’t, and shouldn’t be, a kind of technical school to supply industry with workers, they do have a role to play. Department heads and graduate advisers should be aware of job opportunities in sectors beyond academia. Holding career days, lining up internships and launching seminar series in industry skills could help academia alleviate industry’s woes.

Industry, which often bemoans the lack of qualified candidates, could do more too. Companies could visit various departments and tell faculty members and students alike what would constitute their ideal candidate. They could sponsor more internships and actively promote them. They could initiate research collaborations and involve graduate students in them. And they could offer industrial postdocs as a way to train up potential employees.

Young scientists in training aren’t off the hook, either. They could take more control of their careers by exploring different functions in industry. Taking short courses, doing internships at companies, collaborating with industrial researchers and attending seminars on off-the-bench skills could all provide experience that would be applicable to any sector and put them in contention for industrial jobs.

Whether, when or how the UK government sketches out the road to reform, it’s safe to say that universities, students and companies can embark on it sooner and get to the destination faster.

Paul Smaglik, Naturejobs editor

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestranguez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

Masters of efficiency

Manufacturing jobs may be shifting from the large drug companies to contract organizations as firms re-evaluate their strengths. But scientists with analytical skills and an eye for efficiency can find a job transforming materials into medicines, says **Hannah Hoag**.

The story of a pharmaceutical reads a bit like a fairy tale. There's good and evil (the medicine and the disease), the battle between the two, and — if the author has any hopes of being published — a happy ending. But somewhere along the way, the story of how the good arose gets lost in the drama.

In the pharmaceutical industry, amid the glamour of novel mechanisms and blockbuster drugs, manufacturing is often perceived as dull and devoid of science. This misconception can send scientists joining the industry to jobs in the earlier stages of the drug pipeline before they've even considered manufacturing. Scientists with a gift for tackling scale-up issues, an eye for quality or a talent for squeezing efficiency from a highly regulated process may find that manufacturing offers great career opportunities.

"Many of the same skills are required in pharmaceutical manufacturing as in R&D," says Ron Tull, vice-president of human resources for global manufacturing and supply at GlaxoSmithKline (GSK) in Research Triangle Park, North Carolina. "Science helps us understand our product, how we formulate it, how we granulate it, how we blend it, how we coat it. It's a critical ingredient," he says.

Industry experts are seeing changes in the job market for drug manufacturing. Although career opportunities in manufacturing remain stable, their distribution may be shifting. With big drug firms focusing on their core strengths and their ability to cut costs, some companies are relying on contract manufacturing organizations to meet the needs of their pipeline, whereas others continue to hire, says Rich Pennock, regional director of Kelly Scientific Resources. Amgen, for example, had more than 300 manufacturing positions open across the United States in late July. "Everyday R&D is finding new proteins or new molecules to put through trials. It's not just Amgen, it's all over the industry," says Tony Jacobs, associate director for manufacturing staffing at Amgen in Thousand Oaks, California.

Investment in Europe is also growing. Many

companies, including Wyeth and Altana Pharma, are establishing facilities in Ireland, one of the world's drug-manufacturing hubs. And the newly built National Biomanufacturing Centre in Liverpool, UK, will offer biotech firms the expertise and facilities to manufacture biopharmaceuticals for early-phase clinical trials. The facility could inject as many as 1,000 jobs into the area over the next 5–10 years. Central and eastern Europe have also been pegged as promising — particularly for generics manufacturing — attracting such companies as GSK, Servier and Sanofi-Aventis.

Made to measure

Within a company, manufacturing has two opposing key performance indicators. One calls for efficiency, whereas the other demands the flexibility to react to last-minute marketing forecasts. The challenge is to find the right balance. "We need to make sure that there is enough product, based on marketing forecasts, and that we produce it in a timely, flexible and fast way," says Luanne Shock, executive director of human resources, pharmaceutical operations, at Novartis in Suffern, New York. But from the outside, compliance is key. Like clinical trials (see *Nature* **442**, 480–481; 2006), manufacturing is a highly regulated segment of the drug industry, and those who deviate too far from the script are liabilities.

Broadly speaking, manufacturing delivers products to the market as efficiently as possible. It oversees production, quality control and quality assurance. Team members develop scale-up processes and devise more efficient means of production. Typically, the team gets involved with a drug in the later stages of development, years after its potential was first identified by discovery scientists (see *Nature* **439**, 886–887; 2006).

Production scientists transform raw materials into capsules, tablets and solutions. Scientists can find entry-level positions in the process scale-up area. At Roche's manufacturing facility in Nutley, New Jersey, scientists transform batches weighing less than 10 kilograms to a full batch size, which can hit 900 kilograms.

Luanne Shock: the process must be fast and flexible. Geoffrey Murgatroyd: do you like playing with cars or lawnmowers?



A LEAN SWEEP

Cars may seem no more like pharmaceuticals than aeroplanes are like oranges, but from a manufacturing perspective the two are intrinsically tied.

Henry Ford's principles of efficient manufacturing, first put into place by Toyota, lie at the core of today's pharmaceutical manufacturing business. As drug companies face increasing pressure to cut costs yet maintain the safety and quality of their complex products, they are taking lessons from the car industry to become 'lean' manufacturing machines.

"Lean manufacturing is sweeping the industry," says Angela Roberts of AstraZeneca. As a result of this newer trend, senior scientists, in addition to supervising associates, may find themselves re-engineering processes to cut inventory and cycle time, and reduce waste.

"We're borrowing from the automotive

industry and the electronics industry, which have been challenged to find more efficient means," says Luanne Shock of Novartis.

In 2002, the US Food and Drug Administration (FDA) launched an initiative to bring twenty-first-century regulation to the drug industry. It challenged the industry to better understand and control the manufacturing process by implementing process analytical technology, or PAT, which aims to design and use processes that can consistently ensure a predefined quality at the end of the production line.

"The industry is moving towards online testing," says Geoffrey Murgatroyd of Roche. PAT brings machines and analytical techniques to the production line to give continued readouts of the product's quality. Roche is testing near-infrared spectrophotometry to assess the drug content of capsules as they are made.

"In the next five years we'll see a move towards increasing this process analytical technology — the FDA is pushing it," says Murgatroyd.

H.H.



Angela Roberts: lean manufacturing takes over.

The position typically demands a degree in pharmacy, chemistry or chemical engineering. But the biotech arena also values life-sciences graduates and even those who have studied food science, because techniques such as fermentation are common to food and pharmaceutical production. Recent college graduates can frequently secure a production position with little or no experience.

Some of the best scientists for production jobs are those who are process oriented and enjoy machinery. "Years and years ago, we used to ask applicants: 'What are your hobbies?'," says Geoffrey Murgatroyd, vice-president of technical operations at Roche. "Anyone who said they liked to play with lawnmowers and cars had a priority because we knew they had an understanding of how machinery worked."

Quality control and assurance are also vital to the industry. Both positions require meticulous analytical scientists with a strong understanding of the regulatory environment. Quality-control scientists work in microbiology or analytical chemistry labs, and, depending on the facility's organization, test raw materials before they are mixed and the products as they come off the line. As new technologies are introduced, product testing will occur on the production line, perhaps changing the role of the quality-control scientist in the future.

Quality counts

People in quality assurance ensure that drugs are made in a safe and compliant manner, following standard operating procedures and good manufacturing practices. They check batch documentation and may also liaise with the regulatory agencies and review critical paperwork. "People working in our site quality control determine if the product should be released," says Tull. Chemistry, biochemistry, microbiology and biology degrees are popular among this group, and the positions require up to three years' experience. Five years' experience in validation, process scale-up or manufacturing can lead to supervisory roles within manufacturing.

When choosing between candidates, experience and education are important but not always paramount. "The best people are those with an inquisitive mind, who don't accept things at face value, and who keep



Ron Tull: experience helps.

asking the important questions of why that happened," says Murgatroyd.

Angela Roberts, graduate development manager at AstraZeneca in Macclesfield, UK, agrees: "We need people who will constructively challenge what we do and don't take everything for granted. That's how we get continuous improvement."

Precision also ranks high, says Jay Stout, director of process development at Amgen's Enbrel manufacturing facility in West Greenwich, Rhode Island. "They have to be exacting and really have to know their science."

"Experience isn't required, but it definitely helps," says Tull. "If you have experience working for a pharmaceutical company, make sure you highlight it. You will get attention because you have worked within a regulated industry." Internships (see *Nature* 439, 504–505; 2006) and graduate-development programmes are great ways to get experience and learn more about different aspects of manufacturing, and employers often hire their interns after graduation, says Stout. Courtney Griffin, now a process validation engineer at GSK's manufacturing facility in Zebulon, North Carolina, spent three summers as an intern at the company during her undergraduate degree. "It's the best way to get your foot in the door," says Griffin.

The way to get ahead is to understand the cutting-edge technologies and processes being applied, such as lean manufacturing, Six Sigma (a data-drive approach to eliminating defects in a process) and process analytical technology, or PAT (see 'A lean sweep'). Some colleges and technical institutes also offer short-degree and certificate programmes in drug manufacturing.

As the pharmaceutical manufacturing industry adjusts to its current challenges, sharp and curious scientists will fare well if they can sniff out the jobs. Those with significantly more experience (upwards of seven years) are in high demand at many large companies. For those reluctant to begin their career in manufacturing, it's important to think strategically. Manufacturing introduces the more regulated side of the industry and can serve as a stepping-stone into other portions of the industry, including discovery or regulatory affairs. A first job in manufacturing could help you create your own fairy tale.

Hannah Hoag is a freelance writer based in Montreal, Canada.

Web links

Pharmaceutical Researchers and Manufacturers Association of America
www.phrma.org

Canada's Research-Based Pharmaceutical Companies
www.canadapharma.org

International Federation of Pharmaceutical Manufacturers and Associations
www.ifpma.org

MOVERS

**Michael Morgan, chief scientific officer,
Genome Canada, Ottawa, Canada**



2002-06: Consultant for special projects, Wellcome Trust, London, UK

1998-2002: Director, research partnerships and ventures, and chief executive of the Wellcome Trust Genome Campus, Cambridge, UK

1994-98: Chairman of science funding, Wellcome Trust, London, UK

Michael Morgan, the orchestrator of the Wellcome Trust's genomics initiatives and the new chief scientific officer of Genome Canada, is certain that his far-reaching influence wouldn't have been possible had he followed the typical path of a bench scientist.

While studying microbiology at Trinity College, Dublin, Morgan was seduced by biochemistry as a way to study the chemical processes that unite life. After his PhD at the University of Leicester, UK, during which he probed the regulation of carbohydrate metabolism in *Escherichia coli*, Morgan received a Commonwealth Fund scholarship to do a postdoc wherever he liked in the United States. Studying somatic cell genetics and virology there was refreshing, he says, because high-risk, cutting-edge scientific experiments were encouraged.

Back in Britain, Morgan spent several years lecturing at Leicester. But academia had little support from the UK government during the 1980s and he jumped at the chance of involvement with the Wellcome Trust's new cell- and molecular-biology grants programme. Although its initial budget was a fraction of today's, Morgan enjoyed his newfound access to people and money. "I was able to engage across a broader spectrum of science," he says.

As the Wellcome Trust grew, so did his responsibilities. Morgan's most profound career decision was to facilitate Wellcome's involvement with human genome sequencing efforts. In addition to arguing that data from the human genome sequence should be released quickly into the public domain to encourage international cooperation, Morgan helped form the Sanger and Wellcome Trust policies, says Martin Bobrow, the trust's deputy chairman. Those policies became the foundation for the international data-sharing principles. The Wellcome Trust Genome Campus, Sanger Institute and European Bioinformatics Institute followed, as Morgan helped put in place Britain's much-needed genomics infrastructure.

A reorganization of the trust led to his appointment as director of research partnerships and ventures, where he began managing high-profile projects such as the SNPs Consortium and the Structural Genomics Consortium.

Now Morgan is heading for Genome Canada, a non-profit genomic and proteomic research foundation set up in 2000 and running five new research centres. His mission, he says, is to amass the scientific rigour and expertise necessary to carry out an ambitious scientific agenda. ■

Virginia Gewin

SCIENTISTS & SOCIETIES

A bridge from Portugal to the States

For years, Portugal's low number of PhDs and inefficient academic system have inhibited the generation of knowledge, technological development and modernization. In response, Portuguese governments have invested significantly in higher education at the PhD level. They have awarded thousands of fellowships to encourage students to go abroad as part of their postgraduate work. A new generation of students with a more informed vision of economics, politics and science has emerged.

Set up in 1998, the Portuguese American Postgraduate Society (PAPS) now has about 800 members all over the United States, hailing from fields such as biology, economics and the social sciences. Many are now professors and business managers in the United States, Portugal and other countries. PAPS aims to foster connections between Portuguese postgraduates in the United States and their peers in Portuguese academia and industry.

The latest of PAPS' seven annual forums — Fruitful Connections across the Atlantic: Science and Technology Networks — took place at the University of California, Berkeley, in March. Portugal's minister for the environment and secretary of state for science and technology were

there, along with managers from US companies such as Microsoft, Cisco and Genentech.

Researchers such as Alcino Silva, a neuroscientist at the University of California, Los Angeles, advocated strong networking to bring students and postdocs to the best US laboratories. In one collaboration, academics at the University of California, San Francisco, are working with Portuguese biotech entrepreneurs. In another, managers at Cisco are helping find work for Portuguese students. We also expect networks to be formed in other fields.

PAPS promotes 'learning tours' for Portuguese academics and politicians to visit US entrepreneurs. It also urges a change of attitude, for scientific merit and teaching capabilities to be valued above the amount of time spent in the academic system. PAPS has promoted discussions on this topic through a blog and submitted recommendations to the government.

PAPS helps postgraduates act as a bridge between Portugal and the United States. In the process, we hope to help Portugal improve its performance in science and technology for years to come. ■

Tiago Fleming Outeiro is chairman of PAPS and a research fellow at Harvard Medical School.

GRADUATE JOURNAL

Scheduling my defence

As I get closer to defending my thesis research, the greatest challenge, aside from the thesis itself, has yet to be conquered: finding a date and time at which all my committee members can be present. After several years in graduate school scheduling advisory meetings, then committee meetings, and finally my PhD qualifying and comprehensive exams, I have gained valuable experience in how to get five professors to meet at the same time. But it still hasn't been easy.

At my first advisory meetings, no one except for my main adviser used to show up. On his recommendation, I replaced the members who never attended with professors who took their advising duties more seriously. The problem was solved. With this in mind, I carefully selected my thesis committee on the basis not only of their reputations as scientists, but of their purported social skills and ability to handle students. It has made a big difference.

But despite a great and supportive committee, a defence date has still not been set. This time I must coordinate the schedules of a professor in Bermuda, one on safari in Africa, one travelling the silk road in China, one who remains to be located, and luckily for me, one whose office is next to mine. He was there last time I checked five minutes ago! Wish me luck — and thank you whoever invented e-mail. ■

Andreas Andersson is a final-year PhD student in oceanography at the University of Hawaii.

Tick-tock curly-wurly

Wha...?

Gareth Owens

Professor Michelle Tartuffe examined her reflection in the black glass of a fluorescent night-time bus ride. She insisted on taking the bus home, and the White House insisted that she did not. Seventy-three inches from her Chaco sandals to her wild hair, the prof knew that she stood out in any crowd, and when she spoke she had an accent: Haitian, with a hint of the rhythms of France. But English was her first language and she spoke it with elegance and precision. Her diction was clear, her vocabulary extensive and her enunciation always deliberate, and yet...and yet.

Whenever she talked to one of C. M. Kornbluth's Marching Morons — any one of the random, under-educated ferals who seemed to have invaded her intelligent world — the first barely articulated utterance that dribbled out in reply to any attempt at communication was always the same.

Initiating contact with an easy-to-comprehend opening question, she would address the pierced moron usually found lurking vacantly by her stop, saying something like: "Hi, do you know how long 'til the bus comes?"

Or, to the bling-bedecked moron on the table next to her in JavaStar: "Could you pass me the sugar please?" And the reply from all genders and races of moron was always based around the theme of "Wha...?"

This now formed a very basic part of the prof's diagnostic markers for determining intellectual status. If the first reply to a communication was a request to repeat the initial statement, she knew straight away that she was dealing with a moron.

Since the turn of the century the prof had been getting paid a small Washington fortune for thinking aloud. She had published a paper extrapolating from a basic concept of super-intelligence and how to deal with the first contact with super-intelligent races.

Stephen Hawking's assessment of the situation was that given our own experience with colonization, any first contact with more advanced aliens would probably be more like the movie *Independence Day* than Spielberg's *ET*. The prof agreed, which was why she sat on the bus chewing distractedly on the nail of her right thumb.



First contact, when it came, had been unmistakable. The aliens had not spoken only to a few bespectacled computer geeks parcelling out data packets to like-minded geeks. No, first contact had stopped the world and given it a good rocking. Every radio, every television, every terminal and mobile phone had received the transmission. Every screen that could show words did, and everything that could make a sound spoke:

Tick-tock curly-wurly

And that was it. That was the whole of the message. The transmission was non-directional, appearing to arrive from every which way simultaneously, and the best brains on Earth, as well as television presenters, were stumped as to what the words actually meant. It was clear that the message was for humanity as a whole, not just a few elite governmental high-ups, but it was equally clear that humanity as a whole had no idea what was just said.

The president's first instinct was to find a way of broadcasting a reply. The prof was the head of the appropriate think-tank, so suddenly she found herself on the great carpet of the Oval Office, explaining to the secretary of state, who in turn re-explained to the president, that a technology gap of more than about 200 years would render communication difficult — and, if these aliens were super-intelligent, impossible — unless such advanced creatures were prepared to take the time to talk to us the way we do to cats and dogs.

"Look at it this way, Mr President," she said. "First contact with an equal is in essence no different from when you see someone you like at a party. You think

that you would like to get to know them. You then have to formulate a small and pithy first-contact statement that will both pique their interest and elicit an open-ended response allowing further communication."

"So what you are saying, professor, is that this alien message is actually nothing more than a cheesy pick-up line."

The prof smiled. "Actually, sir, although the message may be seen as having an analogous purpose, it's probably

more complicated than that."

"Yes," said the president, looking at his fingers. "Almost everything seems to turn out that way."

"What do you think the message actually means?" the secretary of state asked, eyes sharp and alert. "Is it a McLuhan-esque test of reasoning where the words actually have no meaning, the mere existence of the transmission being its own message?"

The prof shrugged. "As yet I have no real idea. For example, what goes tick-tock?"

"A clock," said the president.

"A bomb," said the secretary of state.

"And our DNA is wrapped around in a double spiral, curly-wurly fashion. It could be that the message is saying something along the lines of 'shame about your genetic time-bomb', or it might be that the whole message is nothing more than someone bending over and clapping their hands together, like when you call a dog over."

"Why don't we just ask them what they meant?" said the president.

"No!" squawked the prof, a look of blind panic on her face. "Our future survival is staked on our reply. We must figure it out, whatever it takes."

She rode the bus home. The stars had tapped Earth on the shoulder. Her reflection in the window held her gaze.

"Tick-tock curly-wurly," she said. "Tick-tock curly-wurly." ■

Gareth Owens speaks nine languages including Sumerian and Dutch, composes music that would give most people nightmares, and claims to be one of the few university-qualified wizards in the world. He is an occasional contributor of fiction to *Odyssey: Adventures in Science* magazine and has recently completed his first novel, which is full of stompy robots and exploding spaceships.

JACEY

Abstractions



FIRST AUTHOR

Studies of the visual system have revealed a lot about how we process simple features of an object such as its shape, colour and movement.

Much less is known about how the brain encodes the meaning, or category, of visual stimuli. On page 85 of this issue, David Freedman, from Harvard Medical School, reports the identification of monkey brain cells that they believe allot meaning to objects. These cells are located in the lateral intraparietal (LIP) area, a brain region that uses visual information to help monkeys make decisions such as to move their eyes to or from an object. Freedman teamed up with John Assad, also at Harvard, to show that LIP neurons can code for a specific category and retain that information during a delay period. The information is stored in the monkeys' short-term memory. Freedman spoke to *Nature* about how the brain assigns meaning to moving visual targets.

How did you come to study categorization?

I'm driven by the desire to answer the question of how the brain makes sense of our visual surroundings. We have all these objects around us — such as tables and chairs — and we know what to do with them, but how? We taught monkeys to watch tiny dots drifting in any one of 12 directions and then categorize those directions into two groups of six.

Why use direction of motion rather than, say, chairs?

Partly because we already know a lot about how the brain processes visual-motion patterns — the anatomy and physiology are well understood. And also because so much of our behaviour involves reacting to events in our visual world. If something is moving towards you, do you grab it or move out of its way? We need to understand motion to cross a busy street or recognize whether someone is waving hello or motioning for us to go away. Even simple animals such as flies can detect motion and respond appropriately.

Will this work shed light on memory loss?

People often tell me they hope I will make a pill to improve memory. That is probably a long way off, but we are at least zooming in on areas that have a role in storing long-term memories. This will hopefully lead to a better understanding of amnesia and Alzheimer's, in which people often show impaired learning, memory and recognition.

You've studied aspects of vision since you were in college — why the fascination?

As a student, I attended a course about how the brain makes sense of sensations. I realized these were questions I had been thinking about for years, such as: "Does your red look like my red?" ■

MAKING THE PAPER

Kathy Cashman

Volcano monitoring heats up with new magma knowledge

Volcanologists still aren't particularly adept at working out exactly when volcanic eruptions will occur, how they happen or how long they are likely to last. Now, Kathy Cashman at the University of Oregon in Eugene and her colleagues have found a geological clue about the heating of magma (see page 76). This could help modellers better describe the several-kilometre journey this molten rock makes to Earth's surface.

How magma rises to the surface depends on a number of factors. Many magmas are water-rich, and as they rise the external pressure decreases and water is lost in the form of gas. This water loss changes the properties of the melt such that crystals form and the magma's viscosity increases.

Such factors are of great interest to Cashman and her long-time collaborator Jon Blundy, a petrologist at the University of Bristol, UK. Last year, they published work tracking the decompression, gas loss and crystallization of magmas. While evaluating their results, they realized that something else was happening. When a crystal forms, it gives off heat. "We were actually just fussing with the samples and this sort of fell out," Cashman says. "We all of a sudden said, 'Oh my God, look at what we're seeing!'"

Like molasses, magma gets runnier as it heats up, Cashman explains. Tracking magma movement and temperature should help researchers better understand volcano-monitoring data. "Although people had theorized, 'Oh yeah, latent heating should be happening,' we were actually able to document it," says Cashman.

"I try to pose my research questions such that the findings can be used to improve both the models and ideas about how to monitor volcanoes," she adds. She and Blundy study volcanic

inclusions, the pockets of melt trapped by crystals deep below Earth's surface. By examining various inclusions trapped at different times during the magma's ascent, they attempted to piece together how magma changes as it works its way to the surface.

The work relied on 25 years' worth of samples from Mount St Helens in Washington and additional samples collected over a period of five years from the Kamchatka Peninsula, a volcanic hotbed in eastern Russia. Cashman and Blundy analysed selected inclusions with an instrument known as an ion microprobe. This measures the water content and the concentration of different elements by blasting away part of the sample and running it through a mass spectrometer. By relating the inclusions to the temperatures at which they were preserved, these researchers discovered not only that a significant amount of heating occurs during this process, but that magma crystallization is a relatively fast phenomenon, taking years rather than centuries.

Although her work is of help to modellers, Cashman much prefers to be out in the field collecting samples rather than crunching numbers on a computer. One of her favourite tasks is retrieving samples from active volcanoes that are percolating (as opposed to exploding) with lava. Pieces of pasty lava roll down the side of the volcanic dome — the mound that is formed when lava is so viscous that it cannot flow out of the volcanic crater. Cashman and her colleagues run over to blocks that have rolled a safe distance from the dome, knock off a piece with a hammer, then plunge it into the nearby snow to cool.

Building on field work and geological analysis, Cashman and Blundy continue to dissect the finer points of magma and eruptions. They now plan to investigate how the ascent, degassing and crystallization of these small 'packets' of magma vary with eruption style — why magma travels quickly or slowly and what makes it explode through the surface or form a dome. They want to know, says Cashman, not only what makes magma come to the surface but what keeps it coming. ■

KEY COLLABORATORS

In 1994, two Boston College chemists with neighbouring offices discussed their respective scientific frustrations. Amir Hoveyda was having trouble identifying cleaner, more efficient catalysts in a timely manner, and Marc Snapper was seeking uses for the peptides he had been studying. They soon agreed to draw on Snapper's synthesis background and Hoveyda's catalysis experience. The relationship quickly evolved.

"Now, it's a pure collaboration in which two synthetic chemists go after complex problems together," says Hoveyda.

Most recently, they found a more environmentally friendly way to synthesize small molecules for therapeutic use by skipping several waste-producing steps (see page 67).

Hoveyda says the synthetic organic chemistry field has traditionally been wary of two principal investigators sharing credit on single papers. "The

culture almost discourages it," he says. Early on, colleagues warned Snapper that working with a more senior researcher could hurt his career.

But their partnership has been fruitful; the two have received joint grants from the National Institutes of Health since 1997 and have published about 20 papers together. "The reason this collaboration has been so successful is that neither of us cares who gets the credit," Hoveyda adds. ■